

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
MESTRADO INTERINSTITUCIONAL EM LINGÜÍSTICA**

ZULMÉA MUNHOZ DA ROCHA TEIXEIRA

A SEMÂNTICA DA PREPOSIÇÃO DE EM SINTAGMAS NOMINAIS
ESTUDO COMPARATIVO BASEADO EM *CORPUS* DO
GÊNERO TEXTUAL RELATÓRIO

Dissertação de Mestrado

**FLORIANÓPOLIS
2006**

Zulméa Munhoz da Rocha Teixeira

A SEMÂNTICA DA PREPOSIÇÃO DE EM SINTAGMAS NOMINAIS
ESTUDO COMPARATIVO BASEADO EM *CORPUS* DO
GÊNERO TEXTUAL RELATÓRIO

Dissertação apresentada ao Curso de Mestrado Interinstitucional em *Lingüística e Ensino: novas tecnologias*, da Universidade Federal de Santa Catarina - UFSC, como requisito para obtenção do título de Mestre.

Orientador: Prof. Marco Antônio Esteves da Rocha, Dr.

FLORIANÓPOLIS
2006

À querida sobrinha Maria Eduarda, minha filhinha do coração.

Agradecimentos

Quero estender meus agradecimentos a todos os que participaram direta ou indiretamente deste trabalho, e em especial:

Ao meu orientador, Prof. Marco Antônio Esteves da Rocha, pela paciência e dedicação a mim dispensadas, para que eu pudesse concluir este trabalho.

Ao Prof. Martins Dagostim, pelas valiosas contribuições durante o desenvolvimento deste trabalho.

Ao Prof. Carlos Eduardo Cantarelli, pelo incentivo e compreensão durante o período do curso.

Ao Prof. Admilson Teixeira Franco, que tão prontamente atendeu a minha solicitação, cedendo o material necessário para a pesquisa.

Ao Prof. Y. Shimizu, pela sua costumeira disposição em querer ajudar com seus conhecimentos.

À Prof^a. Faimara do Rocio Strauss, pelo tempo que dispensou na correção da estrutura física do meu trabalho.

Ao Prof. Silvino Iagher, pela colaboração e atenção a mim dispensadas.

Aos meus pais, meu irmão e minha cunhada, pelo estímulo dado durante o período deste curso.

À amiga Adelaide Strapasson, pela companhia e apoio ao longo destes dois anos.

A todos os professores que colaboraram respondendo ao questionário.

LISTA DE GRÁFICOS

Gráfico 01 - Frequência relativa da preposição <i>de</i> , conforme Quadro 02.....	43
Gráfico 02 - Porcentagens obtidas no enunciado 1.....	64
Gráfico 03 - Porcentagens obtidas no enunciado 2.....	65
Gráfico 04 - Porcentagens obtidas no enunciado 3.....	66
Gráfico 05 - Porcentagens obtidas no enunciado 4.....	67
Gráfico 06 - Porcentagens obtidas no enunciado 5.....	68
Gráfico 07 - Porcentagens obtidas no enunciado 6.....	69
Gráfico 08 - Porcentagens obtidas no enunciado 7.....	70
Gráfico 09 - Porcentagens obtidas no enunciado 8.....	71
Gráfico 10 - Porcentagens obtidas no enunciado 9.....	72
Gráfico 11 - Porcentagens obtidas no enunciado 10.....	73
Gráfico 12 - Porcentagens obtidas no enunciado 11.....	74
Gráfico 13 - Porcentagens obtidas no enunciado 12.....	75
Gráfico 14 - Porcentagens obtidas no enunciado 13.....	76
Gráfico 15 - Porcentagens obtidas no enunciado 14.....	77
Gráfico 16 - Porcentagens obtidas no enunciado 15.....	78

LISTA DE QUADROS

Quadro 01 - Extensão de <i>Corpus</i>	21
Quadro 02 - Relação das dez palavras de maior frequência em cada um dos <i>corpora</i>	42
Quadro 03 - Lista de palavras-chave, apresentando a preposição de	43
Quadro 04 - Resultado da análise.....	46
Quadro 05 - Preposição de com valor semântico <i>matéria</i>	58
Quadro 06 - Preposição de com valor semântico <i>origem</i>	58
Quadro 07 - Preposição de com valor semântico <i>finalidade</i>	59
Quadro 08 - Preposição de com valor semântico <i>grupo ou série a que pertence</i>	59
Quadro 09 - Preposição de com valor semântico <i>grupo ou série a que pertence</i>	59
Quadro 10 - Preposição de com valor semântico <i>matéria</i>	60
Quadro 11 - Preposição de com valor semântico <i>grupo ou série a que pertence</i>	60
Quadro 12 - Preposição de com valor semântico <i>medida</i>	60
Quadro 13 - Preposição de com valor semântico <i>assunto</i>	61
Quadro 14 - Preposição de com valor semântico <i>medida</i>	61

Quadro 15 - Preposição de com valor semântico <i>posse</i>	61
Quadro 16 - Preposição de com valor semântico <i>grupo ou série a que pertence</i>	62
Quadro 17 - Preposição de com valor semântico <i>meio, instrumento ou modo</i>	62
Quadro 18 - Preposição de com valor semântico <i>finalidade</i>	62
Quadro 19 - Preposição de com valor semântico <i>matéria</i>	63
Quadro 20 - Demonstrativo das respostas coincidentes entre os participantes.....	63
Quadro 21 - Comparativo da análise do valor semântico <i>matéria</i>	64
Quadro 22 - Comparativo da análise do valor semântico <i>origem</i>	65
Quadro 23 - Comparativo da análise do valor semântico <i>finalidade</i>	66
Quadro 24 - Comparativo da análise do valor semântico <i>grupo ou série a que pertence</i>	67
Quadro 25 - Comparativo da análise do valor semântico <i>grupo ou série a que pertence</i>	68
Quadro 26 - Comparativo da análise do valor semântico <i>matéria</i>	69
Quadro 27 - Comparativo da análise do valor semântico <i>grupo ou série a que pertence</i>	70
Quadro 28 - Comparativo da análise do valor semântico <i>medida</i>	71

Quadro 29 - Comparativo da análise do valor semântico <i>assunto</i>	72
Quadro 30 - Comparativo da análise do valor semântico <i>medida</i>	73
Quadro 31 - Comparativo da análise do valor semântico <i>posse</i>	74
Quadro 32 - Comparativo da análise do valor semântico <i>grupo ou série a que pertence</i>	75
Quadro 33 - Comparativo da análise do valor semântico <i>meio, ins- trumento ou modo</i>	76
Quadro 34 - Comparativo da análise do valor semântico <i>finalidade</i>	77
Quadro 35 - Comparativo da análise do valor semântico <i>matéria</i>	78
Quadro 36 - Demonstrativo dos resultados.....	81

LISTA DE SIGLAS

ALPAC	Automatic Language Processing Advisory Committee
Cefet-PR	Centro Federal de Educação Tecnológica do Paraná
ENIAC	Eletronic Numerical Integrator and Computer
IBM	International Business Machine
MANIAC	Mathematical Analyser Numerator Integrator and Computer
MIT	Instituto de Tecnologia de Massachusetts
MT	Machine Translation
NILC	Núcleo Interinstitucional de Lingüística Computacional
OUP	Oxford University Press
PC	Personal Computer
PLN	Processamento de Linguagem Natural
SN	Sintagma nominal
SP	Sintagma preposicionado
SV	Sintagma verbal
UNL	Universal Networking Language

RESUMO

Esta dissertação apresenta um comparativo entre um *corpus* de estudo, composto de relatórios de estágio de alunos do curso de Engenharia Industrial Mecânica do Cefet-PR, e um *corpus* de referência, composto de textos compilados pelo Núcleo Interinstitucional de Lingüística Computacional (NILC), a fim de identificar a palavra ou palavras de maior centralidade (freqüência) na caracterização do gênero textual do *corpus* de estudo. A identificação foi realizada através de um programa de análise de *corpus*, que verifica e compara a freqüência relativa de palavras em *corpora* através de um procedimento estatístico. Embora tivesse sido constatado anteriormente, através de um procedimento preparatório que gera uma lista das freqüências de todas as palavras de cada um dos *corpora*, que a preposição **de** era a palavra mais freqüente em ambos os *corpora*, o procedimento estatístico de comparação selecionou **de** como uma das palavras centrais do *corpus* de estudo. Com base neste resultado, foi realizada a análise semântica da referida preposição em sintagmas nominais. Finalmente, os resultados da análise semântica foram comparados com as respostas de professores de Língua Portuguesa obtidos por meio de um questionário.

Palavras-chave: Lingüística de *corpus*; lingüística computacional; gêneros textuais; sintagma nominal; semântica da preposição **de**.

ABSTRACT

This dissertation presents a comparison between a study *corpus* consisting of training reports made by students from the Mechanical Industrial Engineering course at Cefet-PR, and a reference *corpus* consisting of texts from the *Núcleo Interinstitucional de Lingüística Computacional - NILC* in order to identify the word or words with more keyness (frequency) in the characterisation of the text genre in the corpus of study. The identification was made through a concordancing program which verifies and compares the relative frequency of words in *corpora* through a statistical procedure. Although it had been noticed before, through a preparatory procedure that generates a list of the frequencies of all words in each one of the *corpora*, that the preposition **de** was the most frequent word in both *corpora*, the statistical procedure of comparison selected **de** as one of the key words in the *corpus* of study. Based on this result, it were done a semantic analysis of the referred preposition in noun phrases. Finally, the results of the semantic analysis were compared with the answers of the Portuguese teachers from a questionnaire.

Key-words: *Corpus* linguistics; computational linguistics; text genre; noun phrases; semantics of the preposition **de**.

SUMÁRIO

LISTA DE GRÁFICOS.....	05
LISTA DE QUADROS.....	06
LISTA DE SIGLAS.....	09
RESUMO.....	10
ABSTRACT.....	11
1 INTRODUÇÃO.....	13
2 RESENHA DA LITERATURA.....	18
2.1 LINGÜÍSTICA DE CORPUS.....	18
2.1.1 <i>Corpus</i>	22
2.1.1.1 <i>Corpus</i> de Referência.....	23
2.1.1.2 <i>Corpus</i> de Estudo.....	24
2.2 HISTÓRICO DA INFORMÁTICA.....	25
2.2.1 Processamento de Linguagem Natural - PLN.....	27
2.3 SINTAGMA.....	30
2.4 SEMÂNTICA.....	33
3 METODOLOGIA.....	40
3.1 MÉTODO E TÉCNICAS.....	40
3.2 SEMÂNTICA DA PREPOSIÇÃO DE em N de N.....	48
3.3 QUESTIONÁRIO.....	49
4 LEVANTAMENTO E ANÁLISE DOS DADOS.....	50
4.1 ESTUDO SEMÂNTICO DA PREPOSIÇÃO DE REALIZADO PELA PESQUISADORA.....	50
4.2 ESTUDO SEMÂNTICO DA PREPOSIÇÃO DE REALIZADO PELOS PROFESSORES DE LÍNGUA PORTUGUESA.....	58
4.3 COMPARATIVO.....	64
5 CONCLUSÃO.....	80
REFERÊNCIAS.....	84
ANEXOS	

1 INTRODUÇÃO

É forte a ligação entre os gêneros do discurso e a vida cultural e social. Com a invenção da escrita alfabética¹ ampliaram-se os gêneros textuais, os quais foram se desenvolvendo historicamente, situando-se e integrando-se funcionalmente nas culturas em que se desenvolvem e, por existir certa dificuldade em defini-los, são apreciados em seus usos e condicionamentos sócio-pragmáticos sendo, assim, caracterizados como práticas sócio-discursivas.

Segundo Bakhtin (2000, p. 279 e 281) “qualquer enunciado considerado isoladamente é, claro, individual, mas cada esfera de utilização da língua elabora seus *tipos relativamente estáveis* de enunciados, sendo isso que denominamos *gêneros do discurso*”. Distingue-os, também, como primários, que se referem à fala simples, cotidiana, e secundários, que são mais complexos. Afirma, ainda, que “durante o processo de sua formação, os gêneros secundários absorvem e transmutam os gêneros primários (simples) de todas as espécies, que se constituíram em circunstâncias de uma comunicação verbal espontânea”.

Devido à sua diversidade, os gêneros devem ser escolhidos conforme a intenção do enunciador, considerando que todo enunciado é dotado de significado e transmite a intenção do enunciador, bem como seu estilo próprio quando fala ou escreve. O emissor transmite sua mensagem por meio do enunciado endereçado ao receptor utilizando um código comum constituído de signos lingüísticos, em que o significante é a parte material, sensível, e o significado é o elemento abstrato que carrega a idéia, sendo importante considerar que a carga semântica sempre traz a “contaminação subjetiva” do emissor, nem sempre captada pelo re-

¹ estimam os cientistas que a linguagem escrita tem seus primeiros registros no Período Neolítico, ± 4.000 a.C.

ceptor, como também existe a possibilidade de uma mesma palavra ser polissêmica.

Portanto, a necessidade de comunicação dos seres humanos e sua constante interação, pela fala ou pela escrita, é que faz com que os enunciados sejam formados e transmitam o conteúdo do pensamento, considerado objeto do enunciado. Isso é confirmado por Osakabe (*apud* GERALDI, 2002, p. 13), quando diz que “só se pode conceber sua existência enquanto ligada a um processo pelo qual eu e tu se aproximam pelo significado”. Para Bakhtin (2000, p. 290) “o enunciado satisfaz ao seu próprio objeto (ou seja, ao conteúdo do pensamento enunciado) e ao próprio enunciador”.

Porém, existem enunciados que possuem formas específicas e determinadas, mas nem por isso deixam de refletir a individualidade na língua. Isso é confirmado por Bakhtin (2000, p. 283), quando cita que “as condições menos favoráveis para refletir a individualidade na língua são as oferecidas pelos gêneros do discurso que requerem uma forma padronizada, tais como a formulação do documento oficial, da ordem militar, da nota de serviço, etc.”

Considera-se que a língua - instrumento da comunicação - é dinâmica na sua transformação, e quanto mais evolui uma sociedade, o acervo vocabular aumenta em número e em variáveis semânticas, enriquecendo a produção de enunciados e fazendo com que a comunicação se torne mais precisa. O surgimento de novas tecnologias interferindo diretamente na comunicação dos seres humanos é que auxilia no estudo dos gêneros do discurso.

Com relação à linguagem artificial - linguagem de máquina, linguagem de programação - criada pelo ser humano na sociedade moderna (início do século XX), como um arranjo matemático (código binário), a comunicação não pode ser

polissêmica e nem subjetiva, considerando que é “friamente” denotativa por ser um sistema programado. Portanto, ela não é dinâmica, não é produtiva como a linguagem natural. Tendo sido criada pelo ser humano, a composição da linguagem artificial é constituída de elementos iguais ao da linguagem natural; por exemplo: signos, grafemas, morfemas, relação sintagmática, etc. Além de desempenhar o seu papel específico para o qual foi criada, ela também é uma ferramenta eficiente para estudos de textos, pois com a utilização de programas específicos torna-se possível fazer comparações entre os dados do seu *corpus* com os outros gêneros de discurso de linguagem natural.

Segundo Biber *et al.* (1998) “a investigação da linguagem baseada em *corpus* produz conhecimento até então inexistente acerca da linguagem tal qual ela é empregada por um grupo social”. Ainda afirma que “o estudo de grandes amostras de linguagem permite ao pesquisador atingir maior objetividade, confiabilidade, escopo, e representatividade”.

Atualmente, é crescente a utilização de computadores para a realização de análises lingüísticas, facilitadas por meio de programas que são compostos por ferramentas capazes de propiciar e auxiliar na descoberta de novos aspectos da linguagem com maior rapidez. Porém, segundo Sardinha (2004, p. 84), “a parcela de pesquisa lingüística assistida por computador ainda é minoritária”. Portanto, este estudo torna-se relevante no contexto da lingüística computacional, levando em consideração a importância de investigar características de gêneros do discurso específicos utilizando-se meios eletrônicos, para apontar as palavras de maior centralidade em cada gênero, bem como elaborar pesquisas que levam a obter dados significativos para a Língua Portuguesa em si e para o seu ensino, ressaltando, no caso deste estudo, a importância da utilização de recursos

computacionais e comparações de base estatística como ferramentas na investigação dos gêneros textuais e diversidade textual em *corpora*.

A pesquisadora, estimulada por assuntos vistos durante o curso de mestrado, optou pela realização de um trabalho baseado em *corpus*, mais especificamente, por uma abordagem de base em *corpus* para a investigação das características centrais dos gêneros textuais, a qual se baseia na comparação de *corpora* de gêneros textuais específicos, ou mesmo presumidos, a um *corpus* considerado como um *corpus* de referência, por meio de recursos computacionais e estatísticos.

Quanto aos *corpora*, o primeiro é composto por textos compilados pelo Núcleo Interinstitucional de Lingüística Computacional (NILC), o qual foi cedido pelo professor orientador do trabalho; e o segundo, é composto pelos relatórios de estágio de alunos do Curso de Engenharia Industrial Mecânica do Cefet-PR, considerando a facilidade na coleta desses relatórios junto ao coordenador do referido curso que, através de um contato com o professor orientador de estágio, disponibilizou-os em mídia eletrônica.

O objetivo geral desta pesquisa é a análise semântica das palavras de maior centralidade no *corpus* de relatórios de estágio, de modo a caracterizar o gênero textual proposto.

Para atender o objetivo geral é necessário cumprir as seguintes etapas:

- identificar as palavras de maior frequência (centralidade) no *corpus* de estudo, por meio de um procedimento estatístico;
- comparar a frequência relativa das referidas palavras com o *corpus* de referência;

- verificar possíveis variações de natureza semântica da palavra que se apresentou em primeiro lugar de centralidade no *corpus* de estudo, em consequência de sua centralidade e alta frequência em ambos os *corpora*; e
- comparar os resultados obtidos com os provenientes do questionário endereçado aos professores de Língua Portuguesa.

Para discorrer sobre o tema, esta dissertação foi dividida em cinco capítulos distintos: Introdução, Resenha da Literatura, Metodologia, Levantamento e Análise dos Dados, e Conclusão.

O primeiro capítulo, Introdução, abrange considerações sobre gêneros textuais, bem como o foco de estudo desta pesquisa e os seus objetivos. O segundo capítulo, Resenha da Literatura, contém breve apanhado sobre Lingüística de *corpus*, histórico da informática, sintagma e semântica. O terceiro capítulo apresenta a Metodologia utilizada para o levantamento dos dados, isto é, o método, técnicas e procedimentos para o desenvolvimento da pesquisa. O quarto capítulo refere-se ao Levantamento e Análise dos Dados, apresentando a análise realizada pela pesquisadora e o resultado obtido dos questionários respondidos pelos professores de Língua Portuguesa, para estabelecer um comparativo que leve à comprovação confiável dos resultados. O quinto capítulo, Conclusão, após algumas considerações a respeito do trabalho realizado, apresenta os resultados finais e sugestões para futuras pesquisas em torno da semântica das preposições na Língua Portuguesa.

2 RESENHA DA LITERATURA

Para o desenvolvimento da pesquisa a fim de levantar os dados para análise e resultados do trabalho proposto, é preciso conhecer o que alguns especialistas, em áreas de estudo distintas, abordam quando discorrem sobre Lingüística de *corpus*, histórico da informática, sintagma e semântica.

2.1 LINGÜÍSTICA DE CORPUS

Lingüística de Corpus, para Krieger & Finatto (2004, p. 126) são “estudos lingüísticos que tomam por base o reconhecimento de *corpora* textuais e que realizam análises com apoio de ferramentas informatizadas”. Portanto, o *corpus* lingüístico reflete o caráter da situação na qual foi organizado e registrado, apenas como uma amostra da língua, necessitando ser representativo, pois deve conter a soma das características estruturais. Segundo Biber (*apud* SARDINHA, 2000), os traços lingüísticos não ocorrem com a mesma frequência e a variação desta frequência não acontece de forma aleatória.

O pesquisador deve, inicialmente, levantar um número representativo de enunciados constituindo o seu “universo”. A partir deste acervo de enunciados, são separados aqueles que têm relação com o trabalho a ser realizado, para posterior análise estatística e quantificação. Só os selecionados constituem a amostra, que pode ser parcial ou total, de acordo com o interesse do pesquisador.

Há muitas definições para o termo *corpus*, dadas por alguns autores diretamente ligados à Lingüística de *Corpus*, com as quais se torna possível o seu entendimento.

McEnery & Wilson (2001, p. 29) definem *corpus* como: “any collection of more than one text”. E ainda: “the term corpus is simply the Latin for “body”, hence a corpus may be defined as any body of text”.

Hunston (2002, p. 2) diz que:

linguists have always used the word corpus to describe a collection of naturally occurring examples of language, consisting of anything from a few sentences to a set of written texts or tape recordings, which have been collected for linguistic study. More recently the word has been reserved for collections of texts (or parts of text) that are stored and accessed electronically.

Sanchez (*apud* SARDINHA, 2004, p. 18) define *corpus* como:

um conjunto de dados lingüísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso lingüístico ou de alguns de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise.

A possibilidade da realização de análises comparativas de *corpora* é confirmada por Kennedy (1999, p. 4) quando cita que:

a corpus constitutes an empirical basis not only for identifying the elements and structural patterns which make up the systems we use in a language, but also for mapping out our use of these systems. A corpus can be analysed and compared with other corpora or parts of corpora to study variation.

A Lingüística de Corpus, portanto, explora o conjunto de dados lingüísticos que foi coletado com a finalidade de análises lingüísticas, sendo que um *corpus* computadorizado deve conter dados reais, selecionados e representativos. A questão da representatividade está relacionada com a questão da probabilidade e possibilita a verificação da freqüência de aspectos lingüísticos mais e menos comuns nos *corpora*.

Segundo Sardinha (2000, p. 349-350), a Lingüística de Corpus “trabalha dentro de um quadro conceitual formado por uma abordagem empirista e uma visão da linguagem enquanto sistema probabilístico”. Ele acrescenta, ainda, que “a visão da linguagem enquanto sistema probabilístico pressupõe que, embora muitos traços lingüísticos sejam possíveis teoricamente, eles não ocorrem com a mesma freqüência”.

Leech (*apud* SARDINHA, 2004, p. 22) diz que:

o *corpus* possui uma função representativa. A característica mais facilmente associada à representatividade é justamente a extensão do *corpus*, o que significa em termos simples que para ter representatividade, o *corpus* deve ser o maior possível.

Considera-se que um *corpus* menor não possui a representatividade de um *corpus* maior, já que, quanto maior o número de palavras, maior será a chance de ocorrerem palavras de baixa freqüência, isto é, palavras cujo emprego é pouco freqüente na linguagem.²

² Sardinha (2004, p. 24) diz que “em qualquer *corpus*, as formas de freqüência 1 são a maioria”.

Para Sardinha (2000, p. 344),

a extensão do *corpus* comporta três dimensões. A primeira é o número de palavras. A segunda é o número de textos, a qual se aplica a *corpora* de textos específicos. A terceira é o número de gêneros, registros ou tipos textuais. Esta dimensão se aplica a *corpora* variados, criados para representar uma língua como um todo.

O quadro abaixo, extraído de Sardinha (2004, p. 26), mostra a classificação do *corpus* pelo número de palavras:

Quantidade de palavras	Classificação
Menos de 80 mil	Pequeno
80 a 250 mil	Pequeno-Médio
250 mil a 1 milhão	Médio
1 milhão a 10 milhões	Médio-Grande
10 milhões ou mais	Grande

Quadro 01 - Extensão de *Corpus*.

É interessante citar que o *Brown University Standard Corpus of Present-Day American English*, criado em 1964, foi o primeiro *corpus* lingüístico eletrônico, com um milhão de palavras e que, como afirma Sardinha (2004, p. 2), foi o “fato propulsor do desenvolvimento da Lingüística de Corpus”. Vale ressaltar que os textos, na época, foram informatizados por meio de cartões que eram perfurados individualmente.³

Sardinha (2000, p. 356) também considera que:

³ Sardinha destaca que o primeiro *corpus* eletrônico de linguagem falada é atribuído a John McH. Sinclair, composto de 220.000 palavras.

a Lingüística de Corpus não se resume a um conjunto de ferramentas. Assim, se entendermos metodologia como um *modo típico de aplicar um conjunto de pressupostos de caráter teórico*, então a Lingüística de Corpus pode ser entendida como uma metodologia, pois traz consigo algo mais do que simplesmente o instrumental computacional.

O estudo baseado em *corpus* mostra que, por meio da análise automática de textos, tornam-se possíveis novas formas e aplicações de análise lingüística. Conforme Kennedy (1998, p. 204), “corpus-based approaches to language have introduced new dimensions to linguistic description and to various applications by permitting some degree of automatic analysis of text”.

Portanto, a análise de *corpus* auxilia na investigação e classificação de ocorrências verificadas em *corpora* eletrônicos, falados ou escritos, tornando possível a representação dos fenômenos analisados de maneira precisa.

Existem palavras que, conforme o contexto em que são empregadas, apresentam significados diferentes, como é o exemplo dito por Kennedy (1999, p. 206) “words can show differences in use in different contexts. In one context a word like *bank* could refer to the side of a river, while in another, *bank* could be something you do with your money”.

2.1.1 *Corpus*

Para proceder a análise de palavras-chave é necessário um *corpus* de estudo, ou seja, o *corpus* que se pretende descrever, e um *corpus* de referência, ou seja, o *corpus* que será tido como padrão para a realização da análise, e que tem

como finalidade gerar uma norma para que se possa fazer a comparação das frequências do *corpus* de estudo.

2.1.1.1 *Corpus* de Referência

O *corpus* de referência, ou seja, o *corpus* do NILC, é composto de 34.407.864 palavras e foi criado em 1993 com o objetivo de desenvolver, por meio de uma equipe multidisciplinar, um revisor gramatical automático, através de um projeto de parceria entre a Itautec-Philco e a Fundação de Apoio à Física e à Química, do *campus* da USP - São Carlos. Esse projeto foi criado e gerenciado por docentes e pesquisadores do Instituto de Ciências Matemáticas e de Computação e do Instituto de Física de São Carlos, ambos da USP. Conta, atualmente, com a participação de docentes do Departamento de Computação da Universidade Federal de São Carlos, da Faculdade de Letras da Unesp de Araraquara, e do Departamento de Letras da Universidade Federal de São Carlos.

O *corpus* do NILC, considerado representativo da língua, é composto por textos variados, como os científicos, literários e jornalísticos, fornecendo, portanto, subsídios e possibilitando a realização de uma comparação de base em *corpora*.

As atividades do NILC compreendem pesquisas básicas e aplicadas em várias áreas do Processamento Automático de Linguagem Natural (PLN) com destaque para o Projeto ReGra, que colocou no mercado um revisor gramatical para o Português, incorporado ao editor de textos Word, da Microsoft.

Conforme Nunes *et al.* (1999, p. 68-70),

o ReGra é constituído por três módulos principais: 1- estatístico, que realiza uma série de cálculos fornecendo parâmetros físicos de um texto sob análise, como o número total de

parágrafos, sentenças, palavras, caracteres. O componente mais importante desse módulo, entretanto, é o que fornece o índice de legibilidade. 2- mecânico, que detecta erros facilmente identificáveis que não são percebidos por um corretor ortográfico. 3- gramatical: é o mais importante, e o primeiro passo para a elaboração do módulo lingüístico foi o levantamento de erros (ou inadequações) mais comuns entre usuários de nível médio, como secretárias e profissionais de escritório em geral, e alunos cursando o Ensino Médio ou ingressando na universidade.

2.1.1.2 *Corpus* de Estudo

O *corpus* de estudo é composto pelos relatórios de estágio dos alunos do Curso de Engenharia Industrial Mecânica do Cefet-PR. Relatórios, sejam eles individuais ou coletivos, simples ou complexos, parciais ou completos, periódicos ou eventuais, e técnicos, administrativos, econômicos, científicos, possuem uma forma definida e uma função referencial, considerando a importância da informação a ser dada.

Para Martins e Zilberknop (1993, p. 187-188), “relatório é o documento através do qual se expõem os resultados de atividades variadas”. E também, “relatório é um documento que deve ser objetivo, informativo e apresentável”. Ainda, “a extensão do contexto de um relatório varia de acordo com a importância dos fatos relatados” e “a qualidade essencial deve ser a clareza”.

O Estágio Supervisionado é uma disciplina obrigatória nos cursos superiores de graduação e a apresentação do relatório das atividades realizadas pelo estagiário é necessária para a obtenção do certificado de conclusão do curso.

Conforme o Art. 24, Capítulo VI, do Regulamento da Disciplina Estágio dos Cursos Superiores de Graduação do Cefet-PR, o Relatório Final será avaliado com base nos seguintes aspectos:

- a) compatibilidade do trabalho executado com plano de estágio;
- b) qualidade do trabalho e apresentação do relatório; e
- c) capacidade criativa e inovadora demonstrada através do trabalho.

2.2 HISTÓRICO DA INFORMÁTICA

A palavra *informática* vem do italiano ***informazione matematica***, dados processados por uma máquina de computar, o computador.

No texto *História do Computador*, de Helena Sacerdote, (www.informatic.hpg.ig.com.br/historia.htm) é citado que Charles Babbage, em 1822, construiu uma calculadora de seis dígitos e que, em 1833, projetou uma máquina com capacidade para a realização de tarefas de computação com programa, memória, unidade de controle e periféricos de saída, cujos números seriam manipulados por um engenho com engrenagens e rodas, obedecendo às instruções de um operador por meio de cartões perfurados. Na realidade, esta máquina nunca foi construída. Foi só em 1890 que surgiu o primeiro processador de dados, construído por Hermann Hollerith, para contar e tabular o censo americano. Para fabricar esta máquina, Hollerith teve que fundar uma empresa que veio fazer parte da corporação, conhecida hoje como IBM - *International Business Machine*. Esta empresa colocou no mercado, em 1981, o PC - *Personal Computer*.

No entanto, o primeiro computador a processar informações em tempo real, com entrada de dados a partir de fitas perfuradas e saída em tela CRT (monitor de

vídeo) ou na *flexowriter* (espécie de máquina de escrever) foi o WHIRLWIND, do MIT - *Massachussets Institute of Technology* (Instituto de Tecnologia de Massachussets).

Até meados do século XX, os computadores construídos tinham grandes dimensões, como foi o caso do ENIAC - *Eletronic Numerical Integrator and Computer*. Com o surgimento do transistor, porém, inventado em 1947 por Bardeen, Schockley e Brattain, passaram a ter seu tamanho significativamente reduzido. Ainda em 1953, Jay Forrester construiu a memória magnética. Poucos anos mais tarde, em 1957, o matemático Von Neumann contribuiu para a construção de um computador avançado, chamado de MANIAC - *Mathematical Analyser Numerator Integrator and Computer*.

De acordo com Sardinha (2000, p. 327),

nos anos 60, os computadores *mainframe* passaram a equipar centros de pesquisa universitários e foram sendo aproveitados para a pesquisa em linguagem. Com a popularização dos computadores, foi possibilitado o acesso de mais pesquisadores ao processamento de linguagem natural e, concomitantemente, a sofisticação do equipamento permitiu a consecução de tarefas mais complexas, mais eficientemente, sem falar no aumento da capacidade de armazenamento e na introdução de novas mídias (fitas magnéticas, em vez de cartões *hollerith* perfurados, etc.), as quais facilitaram a criação e manutenção de *corpora* em maior número. Com a entrada em cena dos microcomputadores pessoais, nos anos 80, uma nova onda de mudanças aconteceu, como a popularização de *corpora* e de ferramentas de processamento, o que contribuiu decisivamente para o reaparecimento e fortalecimento da pesquisa lingüística baseada em *corpus*.

A idéia da inteligência artificial iniciou-se com Alan Mathison Turing, nascido em 23 de junho de 1912, em Londres, que desde criança demonstrou interesse pela ciência. Em 1935, concluiu o mestrado em *King's College* e, após receber o *Smith's Prize*, em 1936, com a apresentação de um trabalho sobre a Teoria das Probabilidades, passou a se interessar pela área da computação, demonstrando grande interesse em descobrir se o computador tinha a capacidade de pensar (OKA; ROPERTO, 2000).

Turing inventou um teste, conhecido pela primeira vez em 1950, através da publicação do artigo *Computing Machine and Intelligence*, na revista *Mind*, chamado Teste de Turing. Esse teste tinha a intenção de saber da possibilidade de atribuir a noção de inteligência a uma máquina e consistia em submeter um operador a ficar fechado em uma sala para descobrir se duas entidades ocultas que respondiam suas perguntas, feitas indiretamente através do teclado, eram outras pessoas ou uma máquina (PEREIRA; NAGAI, 2004). Isso evidenciaria que a medida de inteligência de uma máquina acaba sendo a sua capacidade de se comunicar em uma língua humana.

2.2.1 Processamento de Linguagem Natural - PLN

Para Sardinha (2000, p. 328), “o Processamento de Linguagem Natural é uma disciplina com laços fortes com a Ciência da Computação e, embora compartilhe vários temas com a Lingüística de Corpus, as duas mantêm-se independentes”.

Allen (1987, p.1) afirma que o objetivo do processamento de linguagem natural é “poder especificar uma teoria da compreensão e da produção da linguagem num tal nível de minúcia que se possa escrever um programa de

computador que entenda e produza linguagem natural”. Se assim for, os estudos da linguagem irão facilitar a comunicação humano-máquina.

A utilização do computador para investigação das línguas naturais teve seu início na década de 50, com a *tradução automática*. Nunes *et al.* (1999, p. 4) citam que em 1949 foram distribuídas 200 cópias do *Weaver Memorandum*, carta que foi escrita pelo vice-presidente da Fundação Rockefeller, Warren Weaver, conhecedor dos trabalhos sobre criptografia computacional, na qual ele convidava universidades e empresas para desenvolverem pesquisas na área da tradução automática, também conhecida como tradução mecanizada ou MT (*Machine Translation*).

Ainda conforme Nunes *et al.* (1999, p. 5), a primeira reunião científica sobre tradução automática ocorreu nos EUA, em 1952, no Instituto de Tecnologia de Massachusetts (MIT) e a primeira demonstração para o grande público ocorreu dois anos depois, na Universidade de Georgetown. A demonstração apresentou um sistema que traduzia, do russo para o inglês, 50 frases selecionadas de um texto sobre Química. O dicionário construído continha 250 palavras e a gramática escrita para o russo possuía apenas seis regras. A referida tradução não apresentou boa qualidade.

As traduções automáticas sofreram diversas críticas e, em 1964, Bar-Hillel, grande conhecedor do tema, incentivou a divulgação do relatório elaborado pelo Comitê Assessor de Processamento Automático das Línguas Naturais (*Automatic Language Processing Advisory Committee - ALPAC*), que continha uma avaliação negativa das pesquisas até então produzidas. Somente a partir da década de 70 é que os trabalhos nessa área passaram a ser desenvolvidos com fundamentação lingüística.

Em 1970, Winograd criou um sistema operacional que passou a ser um marco do PLN: o sistema SHRDLU ou *mundo dos blocos*, para mostrar que a interação humano-máquina poderia ser uma realidade.

Conforme Nunes (1999, p. 7-8), o PLN pode ser utilizado:

1- na manipulação de bases de dados (sistemas de perguntas e respostas), servindo de módulo de comunicação entre o usuário e a base de dados; 2- sistemas tutores, destacando-se nos sistemas inteligentes a capacidade de executar análises morfológicas e sintáticas, formulando frases em diversos idiomas e contextualizando os exemplos por meio de textos e imagens; 3- sistemas que auxiliam na automação de tarefas administrativas; 4- programação automática, que facilita a interação entre o programador e a máquina; 5- sistema de processamento de textos científicos, que armazena dados que podem ser obtidos pelo usuário por meio de perguntas; 6- sistemas especializados, que são projetados com esquemas complexos de decisão, capazes de agrupar fragmentos de informação numa base de dados e sobre ela operar segundo regras de inferência bastante complexas; 7- tradução automática: os sistemas diretos e os transferenciais efetuam a análise sintática da frase da língua de partida e, através de regras de transferência sintática, constroem a representação sintática da frase da língua de chegada. Nos sistemas interlinguais, a língua de partida e a língua de chegada são intermediadas por uma interlíngua, isto é, uma representação abstrata do significado para a qual a língua de partida é “traduzida” e, a partir da qual, a língua de chegada é “gerada”; 8- sistemas acadêmicos que, com a criação do programa MARGIE, executam dois tipos de operações sobre frases: paráfrase e inferência.

Quanto ao SHRDLU, Nunes (1999, p. 7-8), diz que esse sistema

simulava, sob forma de representação gráfica no monitor do computador, o braço de um robô que manipulava um conjunto de blocos sobre a superfície de uma mesa, por meio da interpretação de instruções em inglês digitadas no teclado do computador. No monitor, via-se o braço do robô executando o que lhe era solicitado.

Assim, o Processamento de Linguagem Natural passou a contribuir de maneira significativa para as teorias lingüísticas, possibilitando aos usuários obterem dados para estudos posteriores, como deve ocorrer neste trabalho.

2.3 SINTAGMA

A comunicação entre os seres humanos acontece pela transmissão de um pensamento através da expressão oral ou escrita. Para Bechara (2003, p. 406) “esta unidade lingüística que faz referência a uma experiência comunicada e que deve ser aceita e depreendida cabalmente pelo nosso interlocutor se dá o nome de enunciado ou período”.

Os enunciados, apesar de possuírem formas variadas, apresentam traços comuns, podendo ser declarativo ou enunciativo, interrogativo, imperativo-exortativo, vocativo e exclamativo.

Ainda segundo Bechara (2003, p. 406-407), os enunciados:

- a) são mensagens completas e de acordo com a situação em que se acham falante e ouvinte;

b) são unidades seqüenciais delimitadas por um silêncio precedente a ele e uma pausa final; são proferidos com um contorno melódico particular.

Dentre os tipos de enunciados acima, é na oração que se alicerça a gramática, pois se caracteriza por um verbo reunindo duas unidades significativas: o sujeito e o predicado. No entanto, a frase também é um tipo de enunciado que não apresenta relação prediativa, como os exemplos:

- Depressa!

- Que calor!

Para Lima (2002, p. 232), “frase é uma unidade verbal com sentido completo e caracterizada por entoação típica: um todo significativo, por intermédio do qual o homem exprime seu pensamento e/ou sentimento”.

Koch e Silva (1989, p. 11) dizem que:

costuma-se entender por frase a expressão verbal de um pensamento, ou seja, todo enunciado suficiente por si mesmo para estabelecer comunicação, sendo que toda frase de uma língua consiste em uma organização, uma combinação de elementos lingüísticos agrupados segundo certos princípios, que a caracterizam como uma estrutura.

Os elementos que constituem uma frase, juntos, dão sentido a ela, constituindo, então, o sintagma.

Ainda segundo Koch e Silva (1989, p. 14):

o sintagma consiste num conjunto de elementos que constituem uma unidade significativa dentro da oração e que mantêm entre si relações de dependência e de ordem.

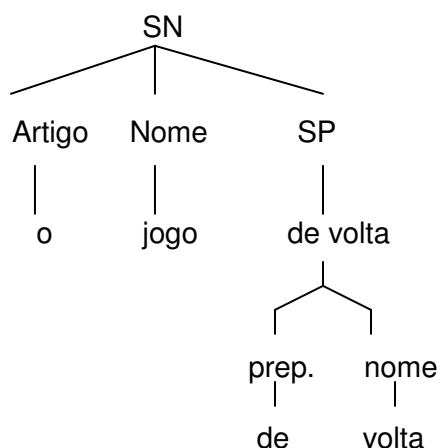
Organizam-se em torno de um elemento fundamental, denominado núcleo, que pode, por si só, constituir o sintagma.

Ferreira (1986, p. 1592) define sintagma como:

a fusão, reunião ou combinação de dois ou mais elementos em que o determinante estabelece um elo de subordinação com o determinado formando ou uma unidade léxica (*vanglória*, em que *vã* é determinante de *glória*) ou locucional (*dona de casa*, em que *de casa* é determinante de *dona*), ou de um termo da oração (*As crianças pequenas choram*, em que os adjuntos adnominais *as* e *pequenas* são determinantes de *crianças*), ou oracional (*O aluno aprendeu a lição*, em que o predicado (*aprendeu a lição*) é determinante do sujeito (*o aluno*)”.

O núcleo é o elemento fundamental do sintagma e o tipo de elemento que o constitui é que vai determinar a natureza do sintagma. Quando esse núcleo é um substantivo (próprio ou comum), um adjetivo ou um pronome substantivo (pessoal, demonstrativo, indefinido, interrogativo, possessivo ou relativo) é chamado de *sintagma* nominal, desempenhando, portanto, uma função na frase. Um substantivo pode ser determinado por artigos, pronomes e numerais e modificado por adjetivos, locuções adjetivas e/ou orações subordinadas adjetivas, sendo que os determinantes antecedem o núcleo e os modificadores são antepostos ou pospostos. O sintagma verbal compreende um verbo, seguido ou não de um sintagma nominal, como pode ser observado no exemplo abaixo:

O jogo de volta	está marcado para 5 de junho, no Mineirão, em Belo Horizonte.
SN	SV



Segundo Cunha e Cintra (1985, p. 120), “podem ocorrer muitos sintagmas nominais (SN) na oração, mas somente um deles será o sujeito”. Afirmam, também, que “a sua posição, na ordem direta e lógica do enunciado, é à esquerda do verbo. Os demais sintagmas nominais encaixam-se no predicado”.

2.4 SEMÂNTICA

Semântica é a parte da gramática que estuda o sentido das palavras de uma língua. Essas palavras podem apresentar vários significados, de acordo com os diversos contextos em que são empregadas. Isso é confirmado por Ilari & Geraldi (2002, p. 5) quando citam que “a semântica é a ciência que estuda a significação”. Moura (2000, p. 63) diz que “calculamos o significado de uma sentença a partir do sentido dessa sentença e da situação em que ela é produzida”.

Com relação ao significado, Saussure diz que “o signo lingüístico une não uma coisa e uma palavra, mas um conceito e uma imagem acústica”. Considerando que signo é a combinação de conceito e de imagem acústica, os quais estão ligados, cabe citar que essa imagem acústica é a representação que se faz de determinada palavra em nossa mente. Portanto, conceito é o significado e imagem

acústica é o significante. Isso é confirmado por Saussure quando diz que “o termo *signo* designa o total e substitui *conceito* e *imagem acústica* respectivamente por *significado* e *significante*”.

Segundo Couto (1983, p. 39-50):

o signo consta de uma expressão e um conteúdo, que são a mesma coisa, ambos são imanentes ao próprio código, nenhum deles transcende, são indissoluvelmente ligados um ao outro. Não há expressão sem conteúdo e não há conteúdo sem expressão.

Para Eco (1973, p. 26-27):

o signo não é apenas um elemento que entra num processo de comunicação (posso também transmitir e comunicar uma série de sons privados de significado), mas é uma entidade que entra num processo de significação.

Portanto, a semântica considera que as palavras e frases estão diretamente ligadas à existência de significados e que a sua interpretação se faz de maneira automática. A palavra *socorro*, por exemplo, pode ter várias significações:

- Socorro é o nome da vizinha.
- Socorro! Estou caindo.
- O Pronto-Socorro atende emergências.

No primeiro exemplo, Socorro designa uma pessoa; no segundo indica que alguém está passando por uma situação de perigo; e o terceiro exemplo refere-se ao local do hospital onde são prestados atendimentos de emergência.

Os elementos componentes das frases, como é o caso das preposições, que não aparecem isoladas e que têm a função única de ligar as palavras, quando

analisadas dentro da frase vão exercer uma função que dará sentido a esta frase podendo, então, apresentar várias significações.

Isso é confirmado por Bechara (2003, p. 297) quando cita que “tudo na língua é semântico, isto é, tudo tem um significado, que varia conforme o papel léxico ou puramente gramatical que as unidades lingüísticas desempenham nos grupos nominais unitários e nas orações”.

Tome-se, como exemplo, um sintagma nominal constituído por um determinado nome (substantivo, adjetivo,...) e um determinante constituído de conectivo vocabular mais substantivo (locução adjetiva):

$$\frac{\text{SN} + (\text{SP} + \text{Nome})}{\text{Subst.} + (\text{Prepositivo} + \text{Subst.})}$$

Acompanhando o sintagma nominal acima, observem-se os dois exemplos a seguir:

a) A pipa ficou presa [.....] telhado.

[no]

[sob]

[sobre]

b) Ele comprou o carro de Pedro.

de Curitiba.

de metal.

O primeiro caso (a) evidencia que a preposição, quando trocada, muda o significado do sintagma nominal “presa....telhado” (que sempre é o mesmo). E, ainda, contamina o significado do enunciado “A pipa ficou presa [...] telhado”.

O segundo caso (b) mostra que o conectivo vocabular não mudou (...de Pedro, ...de Curitiba, ...de metal). O que mudou foram os nomes (determinantes no SP), enquanto o enunciado “Ele comprou o carro de...” se manteve inalterado. No caso (b), a preposição **de** apresentou plurivalência ao trocar o nome que o segue, como se pode observar:

Ele comprou o carro...

de Pedro - idéia de posse (que pertence a Pedro).

de Curitiba - idéia de procedência (que veio de Curitiba).

de metal - idéia de matéria (que foi feita de metal).

Assim, esses dois casos (a e b) mostram que a preposição, além de ser elemento de conexão intervocabular, é portadora de carga semântica, interferindo no significado do sintagma nominal e, conseqüentemente, no enunciado, assunto este que a pesquisa buscará confirmar.

A palavra **de**, por exemplo, que para Almeida (1998, p. 335), “é classificada como essencial na língua, aparecendo como preposição e tendo como função ligar palavras podendo sofrer contrações com os artigos definidos, indefinidos (masculino e feminino), pronomes demonstrativos, pessoal e indefinido, e advérbios”, dependendo do contexto lingüístico em que é empregada, pode apresentar várias significações.

Ex.: a) leite de mulher (origem - procedência do leite)

b) leite da mulher (origem - procedência do leite)

(destino - para a mulher tomar)

Assim, no primeiro caso (a) há univocidade, enquanto o caso (b) pode apresentar dois significados, salvo a situação do contexto estiver expressa, como por exemplo:

- Traz o leite da mulher para a criança tomar.

Gramáticos da Língua Portuguesa mostram que a preposição **de** apresenta valores semânticos distintos, conforme o contexto em que ela aparece. Como observado em Bechara (2003, p. 312-314) a preposição **de** pode indicar:

circunstância de lugar, indicando, origem:

- (1) Sancionada a virtude só pela opinião pública, ela desaparece da vida doméstica e **de** todos aqueles lugares não vistos da multidão.

pessoa, coisa, grupo ou série a que pertence

- (2) A credulidade e confiança **de** muitos tolos faz o triunfo **de** poucos velhacos.

matéria de que uma coisa é feita

- (3) ...ela só lhe aceitava sem relutância os mimos de escasso preço, como a cruz **de** ouro, que lhe deu, uma vez, de festas.

razão ou causa por que uma coisa sucede

- (4) O luxo, como o fogo, devora tudo e perece **de** faminto.

assunto ou objeto de que se trata

- (5) Dizer-se **de** um homem que tem juízo é o maior elogio que se lhe pode fazer.

meio, instrumento ou modo

- (6) O espírito vive **de** ficções, como o corpo se nutre **de** alimentos.

posição, lugar

- (7) Sucede freqüentes vezes admirarmos **de** longe o que **de** perto desprezamos.

medida

- (8) Copo **de** leite.

tempo

(9) **De** noite todos os gatos são pardos.

pode indicar o todo depois de palavras que significam parte

(10) A maioria dos homens, um terço dos soldados, um punhado **de** bravos.

um pouco (ou uma pouca) **de** água.

o modo de ser, semelhança

(11) Muitos figuram **de** Diógenes, para se consolarem de não poderem ser Alexandre.

Em Cunha & Cintra (1985, p. 554-555), observou-se que a preposição **de** pode indicar:

movimento no espaço

(12) Vinha **de** longe, dos confins do medo....,

Movimento no tempo

(13) Como pudera desaparecer **de** um momento para outro, na noção.

Em Rocha Lima (1968, p. 259-260), a preposição **de** pode marcar relação de:

modo

(14) Olhei-o **de** esguelha.

causa

(15) Ninguém morre **de** amor.

origem, procedência

(16) É filho **de** um povoado pobre.

Em Faraco & Moura (1990, p. 290), observou-se que a preposição **de** pode indicar:

assunto

- (17) Vocês podem entender **de** política, **de** economia e **de** engenharia mas **de** cavalos entendo eu... (Senhor).

causa ou motivo

- (18) Leio no jornal a notícia de que um homem morreu **de** fome - F. Sabino.

Também cita que pode indicar:

concessão

- (19) Apesar **de** você, amanhã há de ser outro dia - C. B. Holanda).

matéria

- (20) Vintém **de** cobre ainda o vejo... - C. Coralina.

- (21) Só bebi refresco **de** limão - E. Veríssimo

posse

- (22) O filho **de** Pedro Terra apertou os olhos e a voz - E. Veríssimo.

Para Macambira (1973, p. 66), a preposição **de** pode:

denotar a circunstância de causa

- (23) Não pude comer **de** cansado. A fruta caiu **de** madura.

3 METODOLOGIA

O método utilizado foi o indutivo, através de pesquisa exploratória, quantitativa e qualitativa que, neste estudo, identificou as palavras de maior centralidade nos relatórios de estágio para analisar semanticamente aquela que se apresentou em primeiro lugar de centralidade, estabelecendo as possíveis variações semânticas dentro do sintagma nominal.

3.1 MÉTODO E TÉCNICAS

Considerando a citação de Sardinha (2004, p. 100) de que “o *corpus* de referência não deve conter o *corpus* de estudo, pelo menos não deliberadamente e por completo”, o trabalho teve como ponto de partida a análise de dois corpora: o *corpus* do NILC, considerado de referência, e o *corpus* de relatórios de alunos do Cefet-PR, a fim de estabelecer quais são as palavras de maior centralidade neste último, considerando as diferenças entre as frequências relativas de cada uma das palavras em cada um dos *corpora*. As diferenças são consideradas significativas sempre que a centralidade, conforme medida pelo teste estatístico embutido na ferramenta computacional, ultrapassar um determinado limite.

O primeiro *corpus*, o de referência, ou seja, do NILC, como já citado, é representativo da língua. É composto por textos escritos de diversos gêneros do Português brasileiro, com 34.407.864 palavras.

O segundo, o *corpus* de estudo, proveniente dos relatórios de estágio de alunos do Cefet-PR, é composto por 65.843 palavras. Foi possível a coleta de um número razoável desses relatórios para compor o referido *corpus* de estudo (p. 24)

e, tendo em vista os diversos cursos ofertados pela Instituição, houve a necessidade da escolha de um deles. Em virtude disso, optou-se pelos relatórios de estágio dos alunos do curso de Engenharia Industrial Mecânica, pelo fato de o Coordenador do curso ter providenciado o material em mídia eletrônica.

Para a realização desta pesquisa utilizou-se o programa de análise de *corpus* WordSmith Tools, criado por Mike Scott e publicado há mais ou menos cinco anos pela Oxford University Press, capaz de verificar as palavras de maior centralidade no *corpus* de estudo, ou seja, nos relatórios de estágio.

O referido programa é composto pelas ferramentas WordList - Lista de palavras, KeyWords - Palavras-chave, e Concord - Concordância de palavras, constituídas de instrumentos que propiciam a realização de análises lingüísticas.

Esta exploração dos dados por meios estatísticos acabou por resultar na escolha da preposição **de** como foco principal da pesquisa, tendo em vista a sua centralidade no *corpus* de estudo, conseqüência de seu uso intensivo em sintagmas nominais, como será apresentado na descrição dos resultados (capítulo 4). As variações semânticas observadas neste uso foram então sistematizadas em busca de padrões úteis na caracterização do gênero textual. Com a ferramenta WordList foi possível obter a relação de todas as palavras utilizadas nos referidos *corpora* e suas respectivas freqüências, apresentadas por ordem alfabética ou de freqüência, bem como verificar estatísticas referentes aos dados das listas, conforme mostrado a seguir, onde são indicadas as dez palavras de maior freqüência em cada um dos *corpora*.

NILC			
N	WORD	FREQ.	% LEMMAS
1	DE	1.589.990	4,62
2	A	1.164.956	3,39
3	O	1.074.561	3,12
4	E	776.732	2,26
5	QUE	725.835	2,11
6	DO	632.198	1,84
7	DA	558.919	1,62
8	EM	452.114	1,31
9	PARA	361.395	1,05
10	NO	310.461	0,90

RELATÓRIOS			
N	WORD	FREQ.	% LEMMAS
1	DE	4.488	6,82
2	A	1.954	2,97
3	E	1.874	2,85
4	O	1.466	2,23
5	DO	1.374	2,09
6	DA	972	1,48
7	PARA	943	1,43
8	EM	716	1,09
9	QUE	703	1,07
10	COM	654	0,99

Quadro 02 -Relação das dez palavras de maior freqüência em cada um dos *corpora*.

Observadas as dez primeiras palavras de maior centralidade nos *corpora*, foi realizado um comparativo para observar a freqüência dessas palavras. Quando a freqüência de determinadas palavras é significativamente diferente entre o *corpus* de estudo e o *corpus* de referência, é chamada *chave*. As palavras que apresentam freqüência 1, ou seja, que só aparecem uma vez, denominadas *hapax legomena*, consideradas a maior parte do vocabulário de uma língua, também podem ser verificadas. Cabe aqui lembrar, tomando como base Sardinha (2004), que, quando as diferenças entre as freqüências relativas das palavras são estatisticamente significativas, ou seja, quando a freqüência relativa de uma palavra dada é significativamente mais alta no *corpus* de estudo, é considerada positiva. Já nos casos em que a freqüência relativa de uma palavra dada é mais baixa, é considerada negativa.

Conforme mostrado no quadro acima, a freqüência relativa da preposição **de** no *corpus* de relatórios ou *corpus* de estudo é significativamente mais alta do que a do *corpus* do NILC ou *corpus* de referência, segundo um teste de significância denominado razão logarítmica das probabilidades máximas embutido no programa, sendo, portanto, considerada uma palavra-chave.

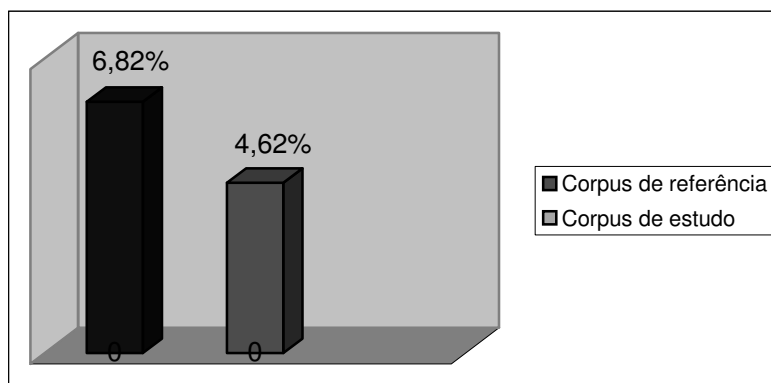


Gráfico 01 - Frequência relativa da preposição **de**, conforme Quadro 02.

Constatou-se, também, que a preposição **de** é uma palavra das palavras mais centrais em relatórios de estágio, confirmando o resultado apresentado pelo NILC. A referida preposição apresentou uma diferença estatisticamente significativa da frequência relativa em cada um dos *corpora*.

Com a utilização da ferramenta KeyWords, gerou-se uma relação de palavras-chave, ou seja, uma lista das palavras que apresentam diferenças estatisticamente significativas nas frequências relativas nos referidos *corpora*, isto é, no *corpus* do NILC e no *corpus* de relatórios.

N	Word	Freq.	NILC %	Freq.	RELATÓRIOS %	Keyness	p
1	ESTÁGIO	987		385	0,58	3.259,7	0,000000
2	MECÂNICA	1.051		141	0,21	925,6	0,000000
3	MONTAGEM	1.638		154	0,23	908,9	0,000000
4	ENGENHARIA	3.400		186	0,28	908,4	0,000000
5	FIGURA	4.870	0,01	205	0,31	900,3	0,000000
6	VOLVO	128		89	0,14	835,5	0,000000
7	EMPRESA	8.591	0,05	292	0,44	755,5	0,000000
8	ATIVIDADES	4.166	0,01	170	0,26	736,6	0,000000
9	FABRICAÇÃO	1.122		112	0,17	673,5	0,000000
10	QUALIDADE	8.073	0,02	195	0,30	655,2	0,000000
11	PROJETO	4.044	0,04	236	0,36	638,6	0,000000
12	COMPONENTES	1.399		110	0,17	612,3	0,000000
13	DE	1.589.990	4,62	4.488	6,82	611,0	0,000000
14	MEDIDOR	66		59	0,09	575,6	0,000000
15	PRODUTO	6.170	0,02	163	0,25	574,5	0,000000

Quadro 03 - Lista de palavras-chave, apresentando a preposição **de**.

As ocorrências foram coletadas e analisadas segundo uma classificação criada especificamente para o propósito. Foram consideradas quatro possíveis categorias de análise para classificar o material segundo o uso da preposição **de**: sintagma nominal do tipo N de N dito técnico; datas; nomes próprios; e locuções, tipicamente associadas ao uso em associação com verbos ou expressões cristalizadas. Mais exemplos podem ser extraídos das concordâncias. Embora o levantamento advenha de textos técnicos, o N de N ocorre em qualquer tipo de texto. Abaixo estão algumas das ocorrências analisadas em sintagma nominal do tipo N de N, tais como:

a) Técnico (codificado **t** na concordância)

- (1) Para este caso, considerado um caso crítico, foram feitos testes com um grande número **de** amostras para avaliar as condições da matéria-prima nacional. (*corpus* de estudo)
- (2) As taxas dos CDBs **de** 30 dias indexados à Taxa Referencial variaram entre 8% e 15% ao ano. (*corpus* do NILC)

b) Datas (codificadas **d** na concordância)

As ocorrências de **de** em datas foram consideradas uma ocorrência especial de sintagma nominal, distinta do tipo N de N dito técnico, e contadas à parte. Exemplos abaixo.

- (3) O grupo Legacy no qual estou trabalhando foi divulgado em um artigo publicado em 04 **de** Fevereiro para toda a Divisão Powertrain no jornal

eletrônico Power News que pode ser acessado através da Intranet da Volvo.

(*corpus* de estudo)

- (4) A Guerra dos Cabanos começou no segundo semestre **de** 1832, no interior de Pernambuco, estendendo-se para Alagoas. (*corpus* do NILC)

c) Nomes próprios (codificados **p** na concordância)

Os nomes próprios também foram considerados um caso especial de sintagma nominal, de modo que as ocorrências desse tipo foram tratadas como uma categoria separada.

- (5) O Complexo Ayrton Senna, da Renault do Brasil, está localizado no Município **de** São José dos Pinhas, região metropolitana de Curitiba, Paraná. (*corpus* de estudo)
- (6) A Secretaria de Transporte **de** Juiz de Fora está realizando uma pesquisa nas plataformas de embarque e desembarque de ônibus na cidade. (*corpus* do NILC)

d) Locuções

Estas ocorrências caracterizam-se por estar associadas a verbos, vinculando a estes últimos os seus complementos. De um modo geral, não constituem sintagma nominal, embora, em alguns casos, expressões consideradas como cristalizadas são contadas como locuções, mesmo tendo aparentemente a estrutura do tipo N de N, como a expressão *de manhã* no exemplo 9 abaixo.

- (7) Encontra-se prevista para meados de abril de 2003, a realização de um lote piloto de 5000 medidores AUDI, a fim **de** se confirmar a eficácia das implementações feitas, assim como, para validar em definitivo o desempenho deste novo produto. (*corpus* de estudo)
- (8) Nas conversas com usineiros, fazendeiros e comerciantes, as críticas de Lula a Collor têm sido acompanhadas **de** aplausos. (*corpus* do NILC)
- (9) Ontem **de** manhã, as pessoas esperavam sob chuva para receber os mantimentos. O comitê distribuiu 300 cestas às famílias mais pobres. (*corpus* do NILC)

Foram analisadas 257 ocorrências extraídas de uma amostra aleatória do *corpus* de estudo por meios computacionais (Anexo I). Subseqüentemente, foram extraídas e analisadas 261 ocorrências do *corpus* de referência, o *corpus* do NILC (Anexo II). A diferença de quatro ocorrências se deve a uma maior incidência de ocorrências consideradas inadequadas no *corpus* de estudo, tais como *mão-de-obra* e *dona-de-casa*, as quais foram tratadas como unidades multipalavra, e não como sintagmas nominais do tipo N de N. O quadro abaixo resume os resultados da análise.

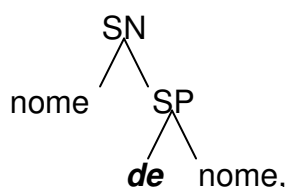
Corpora	N de N	Datas	Nomes próprios	Locuções	Total
Relatórios	190 (73,92%)	10 (8,39%)	12 (4,66%)	45 (17,50%)	257 (100%)
NILC	167 (63,99%)	12 (4,60%)	29 (11,11%)	53 (20,30%)	261 (100%)
Total	357	22	31	98	518

Quadro 04 - Resultado da análise.

Como pode ser observado no quadro, há uma diferença de dez pontos percentuais entre as ocorrências de N de N em cada um dos *corpora*, sinalizando

que a diferença significativa registrada pelo procedimento estatístico de comparação pode ter sido causada, e muito provavelmente o foi, pela maior incidência de sintagmas nominais do tipo N de N, o que não ocorre nas demais categorias. Foi essa a razão do foco de pesquisa ter se deslocado para este tipo de ocorrência. Há muitas outras informações a serem extraídas das concordâncias anotadas, como, por exemplo, a maior quantidade de N de N que nitidamente constituem termos técnicos no *corpus* de estudo, se comparado ao *corpus* do NILC, onde os sintagmas nominais N de N muitas vezes têm um caráter distinto, como, por exemplo, *obras de ampliação* (NILC), que tem um caráter bem menos técnico do que, por exemplo, *rodas de 6" de diâmetro*.

Com base nisso, elaborou-se uma listagem (Anexo III) de enunciados retirados aleatoriamente dos *corpora*, que contivessem a referida preposição na estrutura do sintagma nominal, conforme o seguinte:



ou seja, NOME **de** NOME.

O sintagma nominal do tipo N de N teve uma influência decisiva na seleção da preposição **de** como palavra-chave na comparação entre o *corpus* de estudo e o *corpus* de referência. De um modo geral, as palavras gramaticais não são selecionadas pelos procedimentos estatísticos de comparação de *corpora*. As diferenças entre *corpora* geralmente se manifestam nas palavras ditas lexicais, cuja frequência relativa varia consideravelmente segundo o gênero textual, uma vez que as mudanças no material semântico se refletem na escolha de palavras.

Este efeito pode ser facilmente observado no resultado gerado pelo procedimento de comparação, baseado na razão das probabilidades máximas, que lista palavras como *estágio*, *projeto*, *mecânica* e *componentes* (Quadro 03, p.43), as quais caracterizam o subgênero textual relatórios de estágio em Mecânica. Porém, a preposição **de** também foi selecionada como um dos itens lexicais, cuja centralidade fica demonstrada por uma frequência relativa significativamente diferente, se comparada à registrada no *corpus* de referência. Por que uma palavra gramatical, que repetidamente aparece como a mais freqüente em um grande número de *corpora*, inclusive os dois analisados nesta pesquisa, seria selecionada como uma palavra-chave? Do ponto de vista semântico, parece difícil explicar uma diferença significativa no uso de uma palavra gramatical em um *corpus* de textos de um gênero textual bastante específico. Os números confirmam isso, uma vez que é a única palavra gramatical entre as 15 mais centrais, conforme determinado pelo procedimento.

A hipótese de explicação seria então sustentada por uma maior incidência de sintagmas nominais do tipo N de N, característicos de termos técnicos, os quais provavelmente seriam mais comuns no *corpus* de estudo, a ponto de alterar a frequência relativa da palavra o bastante para tornar a diferença significativa. Uma maneira de tentar demonstrar isso é coletar duas amostras aleatórias de dimensões semelhantes, contendo ocorrências da preposição **de**, em cada um dos *corpora*.

3.2 SEMÂNTICA DA PREPOSIÇÃO **DE** EM N de N

Sendo a preposição **de** uma palavra central no *corpus* dos relatórios, o seguinte passo foi estudar as significações. Uma vez conhecido o valor semântico

das ocorrências da preposição **de** no sintagma nominal, foi realizado um comparativo às respostas de profissionais do ensino, obtidas por meio de um questionário.

As alternativas constantes do referido questionário apresentam variáveis semânticas que tomaram como base gramáticas da Língua Portuguesa, conforme mostrado no item 2.4.

3.3 QUESTIONÁRIO

O questionário (Anexo IV) foi elaborado após estudo das significações da preposição **de**, conforme item 3.2, e é composto por enunciados extraídos dos *corpora*, apresentando ocorrências da preposição **de** no sintagma nominal. O referido questionário deve ser respondido por 20 (vinte) profissionais que ministram aulas de Língua Portuguesa e Literatura Brasileira no Cefet-PR. Na seqüência, deve ser feito um comparativo com os resultados obtidos pela pesquisadora, os quais serão estudados e analisados estatisticamente com o intuito de averiguar a consciência destas funções semânticas em profissionais de ensino.

Como foi dito anteriormente, os enunciados relacionados no questionário provêm da escolha aleatória dos *corpora*, sendo transcritos *ipsis litteris*, *ipsis verbis*, conforme foram redigidos, podendo, portanto, haver erros de linguagem.

4 LEVANTAMENTO E ANÁLISE DOS DADOS

Com este estudo observou-se que, embora houvesse palavras lexicais específicas do gênero textual com frequência alta, a preposição **de** foi selecionada pelo procedimento estatístico como sendo palavra-chave no *corpus* de estudo com base em uma diferença significativa entre as frequências relativas da palavra em cada um dos corpora. A seguir é mostrado o estudo semântico da referida preposição, realizado pela pesquisadora e complementando com o resultado do questionário, remetido a 20 profissionais de Língua Portuguesa, dos quais somente 16 foram os respondentes, correspondendo a 80% de participantes. O comparativo entre os resultados obtidos pela pesquisadora e os fornecidos pelos professores, mostrou as concordâncias e divergências nas variações semânticas da preposição **de**.

4.1 ESTUDO SEMÂNTICO DA PREPOSIÇÃO **DE** REALIZADO PELA PESQUISADORA

O estudo semântico da preposição **de**, em sintagma nominal, realizado pela pesquisadora, apresentou os seguintes significados: 1- *matéria*, 2- *origem*, 3- *finalidade*, 4- *assunto*, 5- *posição*, 6- *medida*, 7- *posse*, 8- *grupo ou série a que pertence*, 9- *causa*, 10- *meio, instrumento ou modo*. Essas significações são exemplificadas através de 15 enunciados extraídos aleatoriamente dos *corpora*. Ressalte-se que, quando num mesmo enunciado a preposição **de** estivesse repetida, seria analisada semanticamente apenas uma dentro daquele enunciado, a qual se apresenta negritada.

Isso, necessariamente, não subentende que são apenas esses os significados em N de N (nome + prep. de + nome). O que a pesquisadora pretende, em virtude da centralidade da preposição **de**, é mostrar a importância das preposições quanto ao seu papel semântico dentro do contexto escrito.

Contudo, sabe-se que é significativa, como se pode ver no exemplo ‘Traga o *leite da vaca*.’..... a idéia é da procedência (de onde é proveniente) do leite; dificilmente pensar-se-á em *leite (para) a vaca tomar*, (o que não deixa de ser possível).

Assim, no estudo dos enunciados selecionados, buscou-se mostrar que a preposição apresenta variações semânticas, entendidas somente dentro de um contexto.

a) Matéria

A preposição **de** com esse significado ocorreu em 2 (dois) enunciados em um total de 15.

(1) O shunt é composto por três partes, sendo duas hastes de cobre e um tarugo cilíndrico de manganina (uma liga de cobre e manganês, cuja resistência elétrica é altamente estável em variações de temperatura). [42]

Neste sintagma nominal, o determinante está transferindo a idéia de matéria através da preposição **de** ao determinado (do que ele é constituído), limitando-lhe, no enunciado acima, o significado: liga feita da matéria cobre; exclui, assim, outros entendimentos. Observe-se que não há o mesmo significado em *liga de (para) meia (finalidade)*, *liga de senhora (posse)*. Se a preposição **de** for substituída por outra, embora sejam mantidas as palavras ‘liga’ e ‘cobre’ do sintagma nominal, o

significado será outro devido tão somente à troca da preposição *liga para cobre* (*finalidade*), *liga com cobre* (*inclusão*), *liga sobre cobre* (*lugar*). Já em ‘liga **em** cobre’ pode-se entender como feita de matéria, portanto a preposição **em** será sinônimo da preposição **de**, ou, ainda, indicando o lugar onde está a liga.

(2) Pão **de** Açúcar renova 50% dos gerentes. [06]

Neste sintagma nominal, exclui-se a possibilidade de ser de outra matéria que não seja ‘açúcar’: pão feito da matéria açúcar. O fato de, neste enunciado, indicar nome de empresa, que por sua vez provém do nome de um acidente geográfico, não elimina o seu significado primeiro, isto é, o sentido denotativo, que seria um pão português adocicado em oposição ao salgado.

Observe-se que outros são os significados em *pão de ontem* (*tempo*), *pão de forno* (*feito onde - lugar ou procedência*), *pão da vovó* (*de quem? procedência*), ou ainda *para quem – destino*), *pão de mistura* (*finalidade*).

b) Origem

A preposição **de** com esse significado ocorreu em 1 (um) enunciado em um total de 15.

(3) Pero Vaz **de** Caminha nasceu, por volta de 1437, provavelmente no Porto.
[117]

Neste caso, a preposição **de** estaria indicando a origem, a procedência de Pero Vaz, fato cultural comum entre alguns povos europeus para indicarem de que

família a pessoa era oriunda, ou, ainda, de que lugar provinha: Pero Vaz é da família Caminha.

c) Finalidade

A preposição **de** com esse significado ocorreu em 5 (cinco) enunciados em um total de 15.

- (4) O período de estágio foi importante ainda na minha formação como profissional, na criação de uma consciência de responsabilidade... [15]

Nesse enunciado, a preposição **de** indica *finalidade*, período destinado a estágio (com o fim de estágio). Mantendo-se o mesmo sintagma, porém substituindo-se o de pela preposição para, *período para estágio*, observa-se que também indica *finalidade*. Neste caso, de e para são preposições sinônimas.

- (5) Programas progressivos de exportação de bens e serviços de informática. [80]

Neste sintagma nominal está explícita a idéia de *finalidade*; substituindo-se a preposição de por para, *programas para exportação*, também têm-se a idéia de *finalidade*. Porém, o significado é outro em *programas sobre exportação (assunto)*, excluindo-se a possibilidade de indicar *lugar*.

- (6) Para isso, será necessário projetar o componente e confeccionar novo molde para o processo de injeção de alumínio sob pressão. [197]

Neste caso, a preposição **de** está indicando finalidade. Alterando-se a preposição de por para, *processo para injeção*, continua com o significado de *finalidade*. No caso de haver substituição da preposição **de** pela preposição com, *processo com injeção*, a idéia é de *modo*.

(7) Mercado ganha nova trava de volante. [313]

Neste enunciado a preposição **de** também indica *finalidade*, ou seja, trava destinada a volante, sendo possível indicar a mesma coisa substituindo a preposição **de** pela preposição para. Observe-se que não existe o mesmo significado em *trava de ferro*, que indica *matéria*.

(8) ... técnicas da empresa transformadora ou de injeção das peças plásticas (conhecedores do processo de injeção de plásticos) e da ferramentaria responsável pela confecção dos moldes... [106]

Neste enunciado, a preposição **de** indica *finalidade*, isto é, processo de injeção destinado, com o fim de injetar plástico. Caso mantiver o mesmo sintagma nominal substituindo-se a preposição *injeção em plástico*, indica *matéria*, *injeção sobre plástico* indica *lugar*, *injeção com plástico*, indica *inclusão*.

d) Assunto ou objeto de que se trata

A preposição **de** com esse significado ocorreu em 2 (dois) enunciados em um total de 15.

- (9) O trabalho realizado durante o período de estágio foi na Fábrica de Veículos de Passeio (CVP), na Engenharia de Montagem no setor de Projetos, na definição e na... [104]

Neste enunciado, a preposição **de** indica *assunto*, isto é, setor que trata de projetos. Se a preposição fosse substituída por para, *setor para projetos* indicaria *finalidade*, isto é, setor destinado a projetos.

- (10) Fome de amor é o Vidas Secas dos ricos. [24]

No sintagma nominal deste enunciado, observa-se que a preposição **de** indica o *assunto*, excluindo outros entendimentos.

e) Medida

A preposição **de** com esse significado ocorreu em 1 (um) enunciado em um total de 15.

- (11) O DST é uma cifra de tempo para o veículo, é através dele que se calculado DSTR, que é uma medida da razão de performance da manufatura (*sic*). [215]

Neste caso, a preposição **de** parece indicar medida de tempo não específico, e sim genérico. Poderia ser cifra de horas, minutos. E no caso de cifra de distância, seria medida de espaço; cifra de dinheiro, a medida indicada seria de valor, importância. Em todos os casos acima o referencial é genérico. Específico seria se fosse cifra de metros, cifra de cinco reais, cifra de litros.

Observe-se, ainda, que esse enunciado (*sic*) foi mal redigido pelo estagiário, apresentando erro de sintaxe e prejudicando, conseqüentemente, a inteligibilidade.

f) Posse

A preposição **de** com esse significado ocorreu em 1 (um) enunciado em um total de 15.

- (12) A Fórmula 1 inicia hoje, no circuito de Nurburgring, na Alemanha, sua participação em terras européias na temporada 96. [16]

Neste enunciado a preposição **de** indica *posse*: o circuito pertence à cidade de Nurburgring. Mantendo-se o sintagma nominal e trocando-se de por em, o significado da preposição será de *lugar*: *circuito em Nurburgring*. No caso da substituição de Nurburgring por TV, ou seja, *circuito de TV*, ter-se-á idéia de *finalidade* (circuito para quê?). Ainda em circuito de Nurburgring, a leitura completa é de circuito da cidade de Nurburgring, confirmando *posse* (de quem é o circuito?).

g) Grupo ou série a que pertence

A preposição **de** com esse significado ocorreu em 3 (três) enunciados em um total de 15.

- (13) O horário da prova é às 8h30, para as áreas de exatas, artes e comunicação, e às 14h30, para as áreas de humanas e sociais. [105]

Neste sintagma nominal a preposição **de** indica *grupo ou série a que pertence*. Há possibilidade de considerar como *assunto*, pois são áreas que tratam de assuntos pertinentes às ciências. Com a alteração do determinante em *área de serviço* e *área de circulação*, pode indicar *finalidade ou lugar*.

- (14) Este está alocado no departamento de Desenvolvimento de Produto, que ainda é composto por Engenharia de Motores, Engenharia de Transmissão e Compras. [207]

Como no enunciado anterior, a preposição **de** também indica o grupo ou série a que pertence. Se a preposição for alterada *Engenharia para motores*, passa a indicar *finalidade*.

- (15) O baú do dinheiro já foi aberto e, a partir de hoje, 11 das melhores Seleções masculinas de vôlei do planeta passam a procurar o caminho para chegar a ele. [73]

Neste enunciado, a preposição **de** parece indicar *grupo*. Mantendo-se o mesmo sintagma nominal e com a troca da preposição de por para, *seleções masculinas para vôlei*, haverá a idéia de *finalidade* (seleções preparadas para jogar vôlei). Já trocando-se vôlei por São Paulo, *seleções masculinas de São Paulo*, haverá idéia de origem e por que não de posse (pertencem à cidade de São Paulo?).

Enfim, ressaltando-se a importância da preposição na Língua Portuguesa quanto ao seu papel semântico dentro de um contexto, exemplifique-se com este último enunciado: seleções de vôlei. Vôlei não contém a idéia de *grupo*, *finalidade*, *etc.* Assim como seleções também não contém tais idéias, mesmo que se queira inferir a idéia de *grupo*, porque no momento em que se substitui a preposição **de** por para, essa idéia de grupo deixa de existir pelo significado *finalidade*.

Ainda convém lembrar que os quinze enunciados selecionados aleatoriamente, acima analisados, apresentam somente sete dos significados da preposição **de**, citados pela pesquisadora. Não apareceram *posição*, *causa*, *meio*,

instrumento ou modo, mencionados por gramáticos. Em decorrência disso, estas variantes semânticas foram inseridas no questionário dos professores, visando à possibilidade de que algum dos respondentes julgasse mais acertado assinalar uma dessas alternativas, como de fato a seguir se confirmará.

4.2. ESTUDO SEMÂNTICO DA PREPOSIÇÃO **DE** REALIZADO PELOS PROFESSORES DE LÍNGUA PORTUGUESA

O estudo semântico da preposição **de** também foi realizado por professores de Língua Portuguesa, através de questionário proposto (Anexo IV), conforme resultado a seguir.

1. O shunt é composto por três partes, sendo duas hastes de cobre e um tarugo cilíndrico de manganina (uma liga **de** cobre e manganês, cuja resistência elétrica é altamente estável em variações de temperatura). [42]

Opções	Professores
origem (procedência de alguém ou de algo de algum lugar)	
posse (pertencer a alguém ou a alguma coisa)	
matéria (do que é feito)	16
posição (onde está alguém ou alguma coisa)	
causa (por que uma coisa sucede)	
assunto ou objeto de que se trata	
meio, instrumento ou modo	
medida (tempo, espaço, volume, massa,...)	
grupo ou série a que pertence	
finalidade	
Outra. Qual?	

Quadro 05 - Preposição **de** com valor semântico *matéria*.

2. Pero Vaz **de** Caminha nasceu, por volta de 1437, provavelmente no Porto. [117]

Opções	Professores
origem (procedência de alguém ou de algo de algum lugar)	11
posse (pertencer a alguém ou a alguma coisa)	
matéria (do que é feito)	
posição (onde está alguém ou alguma coisa)	
causa (por que uma coisa sucede)	
assunto ou objeto de que se trata	
meio, instrumento ou modo	
medida (tempo, espaço, volume, massa,...)	
grupo ou série a que pertence	4
finalidade	
Outra. Qual? elemento de ligação	1

Quadro 06 - Preposição **de** com valor semântico *origem*.

3. Programas progressivos **de** exportação de bens e serviços de informática. [80]

Opções	Professores
origem (procedência de alguém ou de algo de algum lugar)	
posse (pertencer a alguém ou a alguma coisa)	
matéria (do que é feito)	
posição (onde está alguém ou alguma coisa)	
causa (por que uma coisa sucede)	
assunto ou objeto de que se trata	4
meio, instrumento ou modo	1
medida (tempo, espaço, volume, massa,...)	
grupo ou série a que pertence	
finalidade	11
Outra. Qual?	

Quadro 07 - Preposição **de** com valor semântico *finalidade*.

4. O horário da prova é às 8h30, para as áreas de exatas, artes e comunicação, e às 14h30, para as áreas **de** humanas e sociais. [105]

Opções	Professores
origem (procedência de alguém ou de algo de algum lugar)	
posse (pertencer a alguém ou a alguma coisa)	
matéria (do que é feito)	
posição (onde está alguém ou alguma coisa)	
causa (por que uma coisa sucede)	
assunto ou objeto de que se trata	1
meio, instrumento ou modo	
medida (tempo, espaço, volume, massa,...)	
grupo ou série a que pertence	15
finalidade	
Outra. Qual?	

Quadro 08 - Preposição **de** com valor semântico *grupo ou série a que pertence*.

5. Este está alocado no departamento de Desenvolvimento de Produto, que ainda é composto por Eng^a **de** Motores, Engenharia de Transmissão e Compras. [207]

Opções	Professores
origem (procedência de alguém ou de algo de algum lugar)	
posse (pertencer a alguém ou a alguma coisa)	
matéria (do que é feito)	
posição (onde está alguém ou alguma coisa)	
causa (por que uma coisa sucede)	
assunto ou objeto de que se trata	5
meio, instrumento ou modo	
medida (tempo, espaço, volume, massa,...)	
grupo ou série a que pertence	10
finalidade	
Outra. Qual? Adj.adnominal tipificando a que Eng ^a	1

Quadro 09 - Preposição **de** com valor semântico *grupo ou série a que pertence*.

6. Pão **de** Açúcar renova 50% dos gerentes. [06]

Opções	Professores
origem (procedência de alguém ou de algo de algum lugar)	
posse (pertencer a alguém ou a alguma coisa)	
matéria (do que é feito)	10
posição (onde está alguém ou alguma coisa)	
causa (por que uma coisa sucede)	
assunto ou objeto de que se trata	
meio, instrumento ou modo	
medida (tempo, espaço, volume, massa,...)	
grupo ou série a que pertence	2
finalidade	
Outra. Qual? Conectivo, compl. nome "Pão", Marca, Nome	1, 2, 1

Quadro 10 - Preposição **de** com valor semântico *matéria*.7. O trabalho realizado durante o período de estágio foi na Fábrica de Veículos de Passeio (CVP), na Engenharia de Montagem no setor **de** Projetos, na... [104]

Opções	Professores
origem (procedência de alguém ou de algo de algum lugar)	2
posse (pertencer a alguém ou a alguma coisa)	
matéria (do que é feito)	
posição (onde está alguém ou alguma coisa)	1
causa (por que uma coisa sucede)	
assunto ou objeto de que se trata	3
meio, instrumento ou modo	
medida (tempo, espaço, volume, massa,...)	
grupo ou série a que pertence	8
finalidade	
Outra. Qual? lugar	2

Quadro 11 - Preposição **de** com valor semântico *grupo ou série a que pertence*.8. O período **de** estágio foi importante ainda na minha formação como profissional, na criação de uma consciência de responsabilidade... [15]

Opções	Professores
origem (procedência de alguém ou de algo de algum lugar)	
posse (pertencer a alguém ou a alguma coisa)	
matéria (do que é feito)	
posição (onde está alguém ou alguma coisa)	
causa (por que uma coisa sucede)	
assunto ou objeto de que se trata	3
meio, instrumento ou modo	
medida (tempo, espaço, volume, massa,...)	8
grupo ou série a que pertence	
finalidade	4
Outra. Qual? ?	1

Quadro 12 - Preposição **de** com valor semântico *medida*.

9. Fome **de** amor é o Vidas Secas dos ricos. [24]

Opções	Professores
origem (procedência de alguém ou de algo de algum lugar)	
posse (pertencer a alguém ou a alguma coisa)	
matéria (do que é feito)	1
posição (onde está alguém ou alguma coisa)	
causa (por que uma coisa sucede)	2
assunto ou objeto de que se trata	9
meio, instrumento ou modo	
medida (tempo, espaço, volume, massa,...)	
grupo ou série a que pertence	
finalidade	
Outra. Qual? Definição	2
Tipo, qualidade	2

Quadro 13 - Preposição **de** com valor semântico *assunto*.10. O DST é uma cifra **de** tempo para o veículo, é através dele que se calculado DSTR, que é uma medida da razão de performance da manufatura. [215]

Opções	Professores
origem (procedência de alguém ou de algo de algum lugar)	
posse (pertencer a alguém ou a alguma coisa)	
matéria (do que é feito)	
posição (onde está alguém ou alguma coisa)	
causa (por que uma coisa sucede)	
assunto ou objeto de que se trata	2
meio, instrumento ou modo	4
medida (tempo, espaço, volume, massa,...)	8
grupo ou série a que pertence	1
finalidade	
Outra. Qual? qualificação	1

Quadro 14 - Preposição **de** com valor semântico *medida*.11. A Fórmula 1 inicia hoje, no circuito **de** Nurburgring, na Alemanha, sua participação em terras européias na temporada 96. [16]

Opções	Professores
origem (procedência de alguém ou de algo de algum lugar)	
posse (pertencer a alguém ou a alguma coisa)	12
matéria (do que é feito)	
posição (onde está alguém ou alguma coisa)	1
causa (por que uma coisa sucede)	
assunto ou objeto de que se trata	
meio, instrumento ou modo	
medida (tempo, espaço, volume, massa,...)	
grupo ou série a que pertence	1
finalidade	
Outra. Qual? Espaço	2

Quadro 15 - Preposição **de** com valor semântico *posse*.

12. O baú do dinheiro já foi aberto e, a partir de hoje, 11 das melhores Seleções masculinas **de** vôlei do planeta passam a procurar o caminho.... [73]

Opções	Professores
origem (procedência de alguém ou de algo de algum lugar)	
posse (pertencer a alguém ou a alguma coisa)	
matéria (do que é feito)	
posição (onde está alguém ou alguma coisa)	
causa (por que uma coisa sucede)	
assunto ou objeto de que se trata	1
meio, instrumento ou modo	
medida (tempo, espaço, volume, massa,...)	
grupo ou série a que pertence	15
finalidade	
Outra. Qual?	

Quadro 16 - Preposição **de** com valor semântico *grupo ou série a que pertence*.

13. Para isso, será necessário projetar o componente e confeccionar novo molde para o processo **de** injeção de alumínio sob pressão. [197]

Opções	Professores
origem (procedência de alguém ou de algo de algum lugar)	
posse (pertencer a alguém ou a alguma coisa)	
matéria (do que é feito)	1
posição (onde está alguém ou alguma coisa)	
causa (por que uma coisa sucede)	1
assunto ou objeto de que se trata	2
meio, instrumento ou modo	6
medida (tempo, espaço, volume, massa,...)	
grupo ou série a que pertence	1
finalidade	5
Outra. Qual?	

Quadro 17 - Preposição **de** com valor semântico *meio, instrumento ou modo*.

14. Mercado ganha nova trava **de** volante. [313]

Opções	Professores
origem (procedência de alguém ou de algo de algum lugar)	
posse (pertencer a alguém ou a alguma coisa)	1
matéria (do que é feito)	
posição (onde está alguém ou alguma coisa)	
causa (por que uma coisa sucede)	
assunto ou objeto de que se trata	2
meio, instrumento ou modo	2
medida (tempo, espaço, volume, massa,...)	
grupo ou série a que pertence	
finalidade	10
Outra. Qual? tipificação	1

Quadro 18 - Preposição **de** com valor semântico *finalidade*.

15. ... técnicas da empresa transformadora ou de injeção das peças plásticas (conhecedores do processo de injeção **de** plásticos) e da ferramentaria responsável pela confecção dos moldes... [106]

Opções	Professores
origem (procedência de alguém ou de algo de algum lugar)	
posse (pertencer a alguém ou a alguma coisa)	
matéria (do que é feito)	7
posição (onde está alguém ou alguma coisa)	
causa (por que uma coisa sucede)	
assunto ou objeto de que se trata	4
meio, instrumento ou modo	
medida (tempo, espaço, volume, massa,...)	
grupo ou série a que pertence	3
finalidade	1
Outra. Qual? tipificação	1

Quadro 19 - Preposição **de** com valor semântico *matéria*.

Levantados os dados decorrentes das respostas dos questionários endereçados aos professores de Língua Portuguesa, observa-se que em algumas das situações houve concordância, ocorrendo apenas 01 (um) caso, referente ao primeiro enunciado, em que houve resposta única, conforme mostra o quadro a seguir.

Qtde. de enunciados	Nº do enunciado	Respostas iguais	Percentual
01	01	16	100%
02	04, 12	15	93,75%
01	11	12	75%
02	02, 03	11	68,75%
03	05, 06, 14	10	62,5%
01	09	9	56,25%
03	07, 08, 10	8	50%
01	15	7	43,75%
01	13	6	37,5%
Total: 15	Total: 15		

Quadro 20 - Demonstrativo das respostas coincidentes entre os participantes.

O passo seguinte foi confrontar os resultados do estudo da pesquisadora com as respostas fornecidas nos questionários dos professores para apurar o percentual de equivalência.

4.3 COMPARATIVO

Foi realizado um comparativo entre os resultados obtidos pela pesquisadora com os resultados apresentados pelos professores, cujas porcentagens são mostradas por meio de gráficos. As legendas representam o índice percentual correspondente às respostas dos professores pesquisados.

1. O shunt é composto por três partes, sendo duas hastes de cobre e um tarugo cilíndrico de manganina (uma liga **de** cobre e manganês, cuja resistência elétrica é altamente estável em variações de temperatura). [42]

Opções	Pesquisadora	Professores
origem (procedência de alguém ou de algo de algum lugar)		
posse (pertencer a alguém ou a alguma coisa)		
matéria (do que é feito)	X	16
posição (onde está alguém ou alguma coisa)		
causa (por que uma coisa sucede)		
assunto ou objeto de que se trata		
meio, instrumento ou modo		
medida (tempo, espaço, volume, massa,...)		
grupo ou série a que pertence		
finalidade		
Outra. Qual?		

Quadro 21 - Comparativo da análise do valor semântico *matéria*.

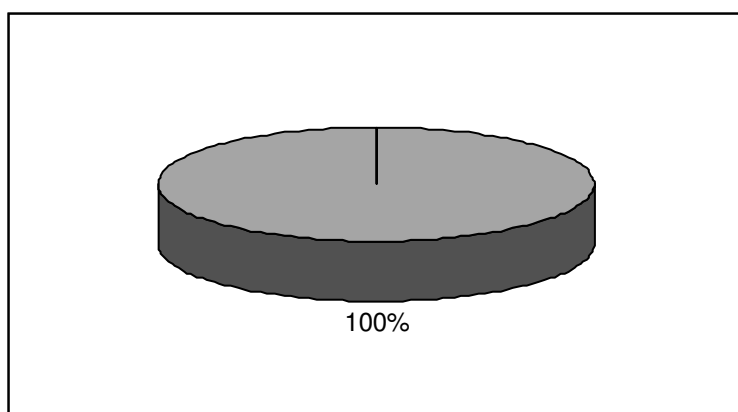


Gráfico 02 - Porcentagens obtidas no enunciado 1

O gráfico acima mostra a equivalência em 100% das respostas dadas pelos professores com o que a pesquisadora concluía quanto ao valor semântico da preposição **de** no enunciado 1, ou seja, que indica *matéria*.

2. Pero Vaz **de** Caminha nasceu, por volta de 1437, provavelmente no Porto. [117]

Opções	Pesquisadora	Professores
origem (procedência de alguém ou de algo de algum lugar)	X	11
posse (pertencer a alguém ou a alguma coisa)		
matéria (do que é feito)		
posição (onde está alguém ou alguma coisa)		
causa (por que uma coisa sucede)		
assunto ou objeto de que se trata		
meio, instrumento ou modo		
medida (tempo, espaço, volume, massa,...)		
grupo ou série a que pertence		4
finalidade		
Outra. Qual? elemento de ligação		1

Quadro 22 - Comparativo da análise do valor semântico *origem*.

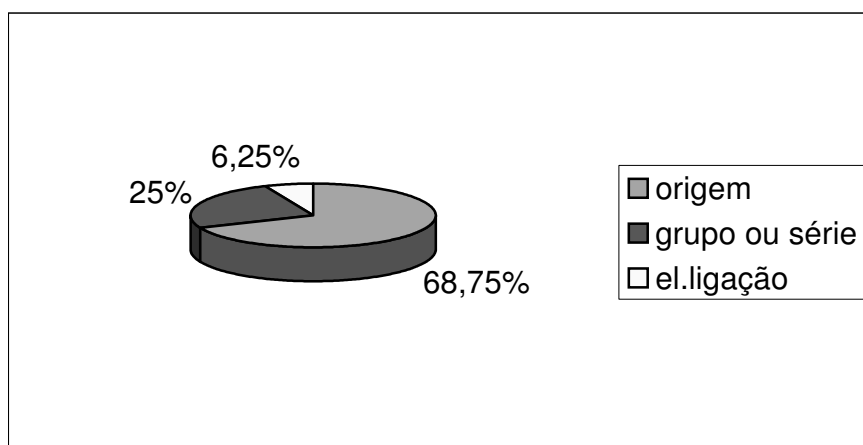


Gráfico 03 - Porcentagens obtidas no enunciado 2.

O resultado da análise realizada pela pesquisadora de que a preposição **de** indica a *origem ou procedência* de alguém foi confirmada em 68,75%. Porém 25% dos professores responderam que indica *grupo ou série a que pertence*. Já 6,25%, ou seja, um caso, respondeu que é um elemento de ligação, resposta que foge ao solicitado: semântica.

3. Programas progressivos **de** exportação de bens e serviços de informática. [80]

Opções	Pesquisadora	Professores
origem (procedência de alguém ou de algo de algum lugar)		
posse (pertencer a alguém ou a alguma coisa)		
matéria (do que é feito)		
posição (onde está alguém ou alguma coisa)		
causa (por que uma coisa sucede)		
assunto ou objeto de que se trata		4
meio, instrumento ou modo		1
medida (tempo, espaço, volume, massa,...)		
grupo ou série a que pertence		
finalidade	X	11
Outra. Qual?		

Quadro 23 - Comparativo da análise do valor semântico *finalidade*.

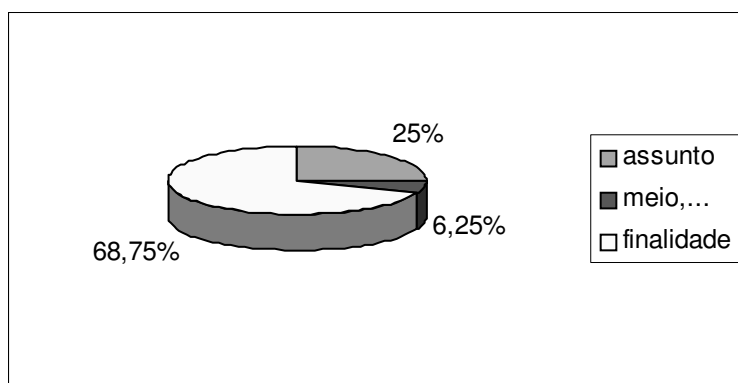


Gráfico 04 - Porcentagens obtidas no enunciado 3.

Neste enunciado, 68,75% confirmaram o resultado obtido pela pesquisadora, ou seja, que a preposição **de** indica a *finalidade*. Outros 25% responderam que indica *assunto*, e 6,25%, isto é, um professor respondeu que indica *meio, instrumento ou modo*, não sendo este último representativo.

4. O horário da prova é às 8h30, para as áreas de exatas, artes e comunicação, e às 14h30, para as áreas **de** humanas e sociais. [105]

Opções	Pesquisadora	Professores
origem (procedência de alguém ou de algo de algum lugar)		
posse (pertencer a alguém ou a alguma coisa)		
matéria (do que é feito)		
posição (onde está alguém ou alguma coisa)		
causa (por que uma coisa sucede)		
assunto ou objeto de que se trata		1
meio, instrumento ou modo		
medida (tempo, espaço, volume, massa,...)		
grupo ou série a que pertence	X	15
finalidade		
Outra. Qual?		

Quadro 24 - Comparativo da análise do valor semântico *grupo ou série a que pertence*.

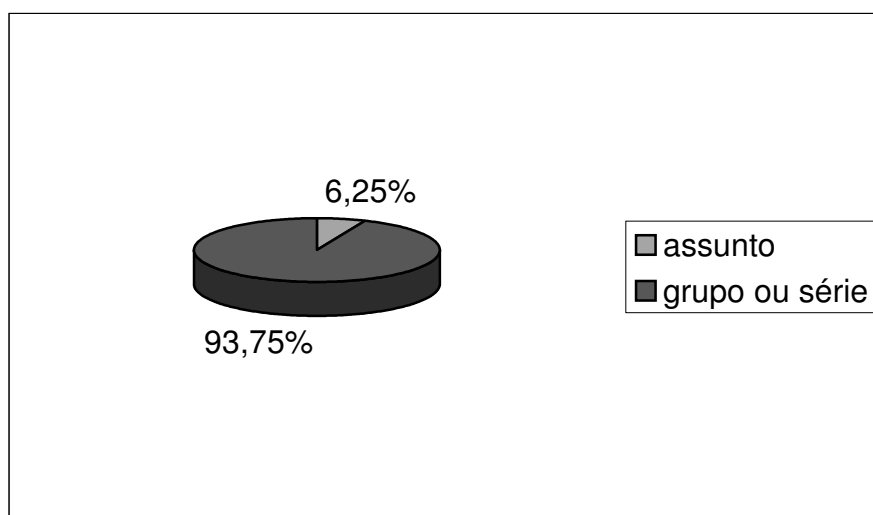


Gráfico 05 - Porcentagens obtidas na frase 4.

Para a pesquisadora, a preposição **de** indica *grupo ou série a que pertence*, sendo que 93,75% comprovaram a referida análise. Porém, 6,25%, ou seja, um professor, respondeu que indica *assunto*.

Há que observar, ainda, que o núcleo do determinante está oculto, ou seja, áreas de ciências humanas (e ciências sociais).

5. Este está alocado no departamento de Desenvolvimento de Produto, que ainda é composto por Engenharia **de** Motores, Engenharia de Transmissão e Compras. [207]

Opções	Pesquisadora	Professores
origem (procedência de alguém ou de algo de algum lugar)		
posse (pertencer a alguém ou a alguma coisa)		
matéria (do que é feito)		
posição (onde está alguém ou alguma coisa)		
causa (por que uma coisa sucede)		
assunto ou objeto de que se trata		5
meio, instrumento ou modo		
medida (tempo, espaço, volume, massa,...)		
grupo ou série a que pertence	X	10
finalidade		
Outra. Qual? Adjunto adnominal tipificando a que Engenharia está se fazendo referência		1

Quadro 25 - Comparativo da análise do valor semântico *grupo ou série a que pertence*.

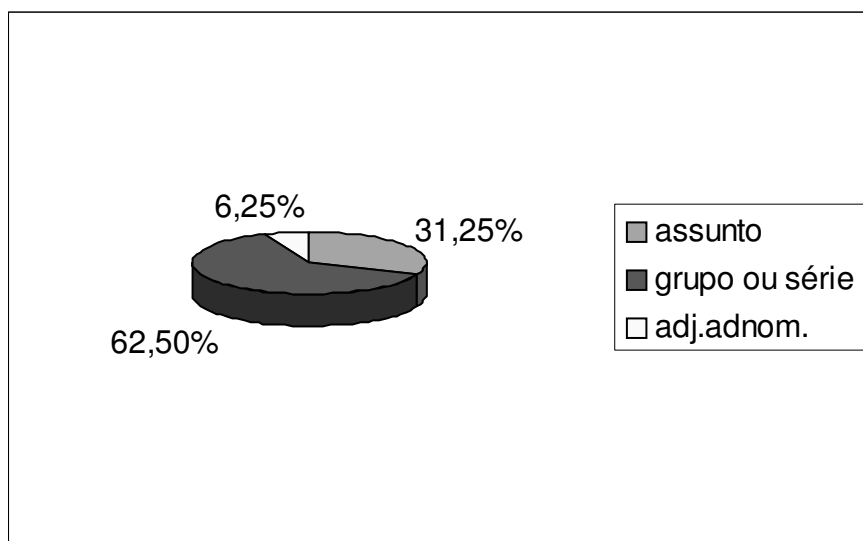


Gráfico 06 - Porcentagens obtidas no enunciado 5.

O gráfico acima mostra que 62,50% concordaram com a pesquisadora de que a preposição **de** indica *grupo ou série a que pertence*. 31,25% responderam que indica *assunto*. Em um caso (6,25%) houve a resposta de que a referida preposição é um adjunto adnominal, resposta que diz respeito à sintaxe e não à semântica.

6. Pão **de** Açúcar renova 50% dos gerentes. [06]

Opções	Pesquisadora	Professores
origem (procedência de alguém ou de algo de algum lugar)		
posse (pertencer a alguém ou a alguma coisa)		
matéria (do que é feito)	X	10
posição (onde está alguém ou alguma coisa)		
causa (por que uma coisa sucede)		
assunto ou objeto de que se trata		
meio, instrumento ou modo		
medida (tempo, espaço, volume, massa,...)		
grupo ou série a que pertence		2
finalidade		
Outra. Qual?		
Conectivo, complemento nome "Pão"		1
Marca		2
nome		1

Quadro 26 - Comparativo da análise do valor semântico *matéria*.

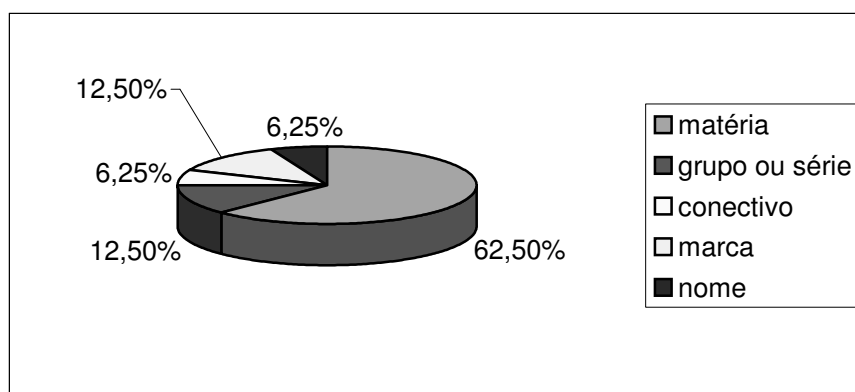


Gráfico 07 - Porcentagens obtidas no enunciado 6.

Neste enunciado, 62,50% confirmam o resultado da análise realizada pela pesquisadora, isto é, *matéria*. Porém, 12,50% responderam que indica *grupo ou série a que pertence*. Outros 12,50% responderam que indica *marca*; um caso, 6,25%, respondeu que é um conectivo. Também um caso, 6,25%, respondeu que indica um nome. Vale ressaltar que 4 respostas, neste enunciado 6, não atenderam ao que foi solicitado, ou seja, o foco era a semântica da preposição **de** e não que ela é um conectivo (1), marca (2) e nome (1). No caso das respostas “marca” e “nome”, os respondentes não se aperceberam que englobavam o sintagma nominal N de N, composto de nome + (preposição **de** e nome).

7. O trabalho realizado durante o período de estágio foi na Fábrica de Veículos de Passeio (CVP), na Engenharia de Montagem no setor **de** Projetos, na definição e na... [104]

Opções	Pesquisadora	Professores
origem (procedência de alguém ou de algo de algum lugar)		2
posse (pertencer a alguém ou a alguma coisa)		
matéria (do que é feito)		
posição (onde está alguém ou alguma coisa)		1
causa (por que uma coisa sucede)		
assunto ou objeto de que se trata	X	3
meio, instrumento ou modo		
medida (tempo, espaço, volume, massa,...)		
grupo ou série a que pertence		8
finalidade		
Outra. Qual? lugar		2

Quadro 27 - Comparativo da análise do valor semântico *grupo ou série a que pertence*.

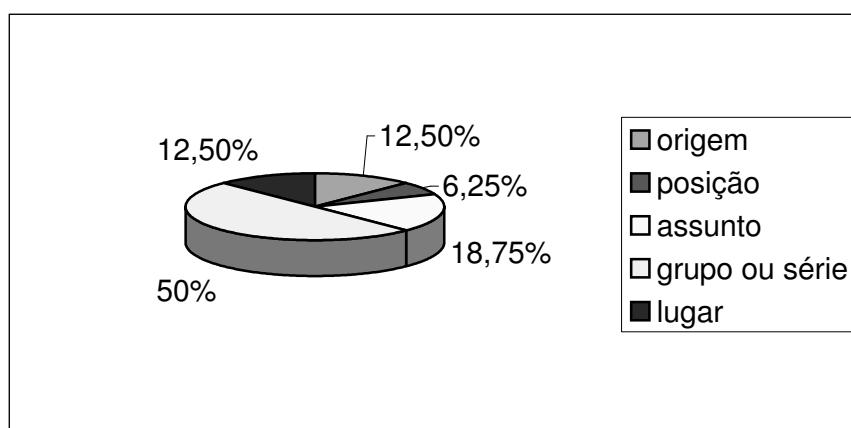


Gráfico 08 - Porcentagens obtidas no enunciado 7.

O resultado obtido pela pesquisadora foi comprovado em 18,75%, ou seja, o **de** neste enunciado 7 indica *assunto*. 50% responderam que a preposição **de** indica *grupo ou série a que pertence*; 12,50% que indica *origem*; outros 12,50% que indica *lugar*; e um caso, 6,25%, que indica *posição*. Há que considerar a discordância dos 50% respondentes em relação aos 18,75% mais a pesquisadora.

8. O período **de** estágio foi importante ainda na minha formação como profissional, na criação de uma consciência de responsabilidade... [15]

Opções	Pesquisadora	Professores
origem (procedência de alguém ou de algo de algum lugar)		
posse (pertencer a alguém ou a alguma coisa)		
matéria (do que é feito)		
posição (onde está alguém ou alguma coisa)		
causa (por que uma coisa sucede)		
assunto ou objeto de que se trata		3
meio, instrumento ou modo		
medida (tempo, espaço, volume, massa,...)		8
grupo ou série a que pertence		
finalidade	X	4
Outra. Qual? ?		1

Quadro 28 - Comparativo da análise do valor semântico *medida*.

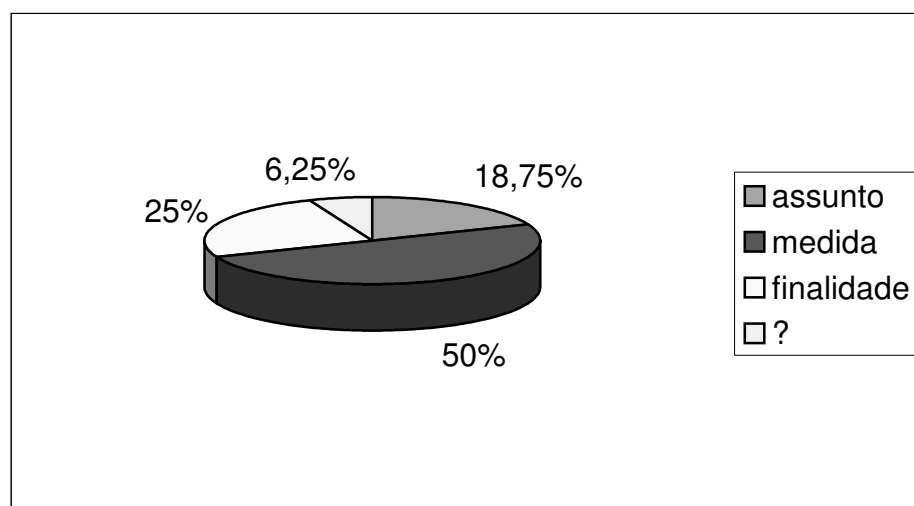


Gráfico 09 - Porcentagens obtidas no enunciado 8.

Foi comprovado em 25% o resultado obtido pela pesquisadora de que indica *finalidade*. 50% responderam que indica *medida*; 18,75% que indica *assunto*. Ainda um caso (6,25%) não disse o que significaria.

Neste enunciado o percentual de respostas que coincidem com a da pesquisadora é inferior (25%) ao da maioria, que também ficou num limite (50%) não convincente. Observe-se, ainda, que na alternativa “outra” e “Qual”? houve uma resposta não especificando qual seria a idéia.

9. Fome **de** amor é o Vidas Secas dos ricos. [24]

Opções	Pesquisadora	Professores
origem (procedência de alguém ou de algo de algum lugar)		
posse (pertencer a alguém ou a alguma coisa)		
matéria (do que é feito)		1
posição (onde está alguém ou alguma coisa)		
causa (por que uma coisa sucede)		2
assunto ou objeto de que se trata	X	9
meio, instrumento ou modo		
medida (tempo, espaço, volume, massa,...)		
grupo ou série a que pertence		
finalidade		
Outra. Qual?		
Definição		2
Tipo, qualidade		2

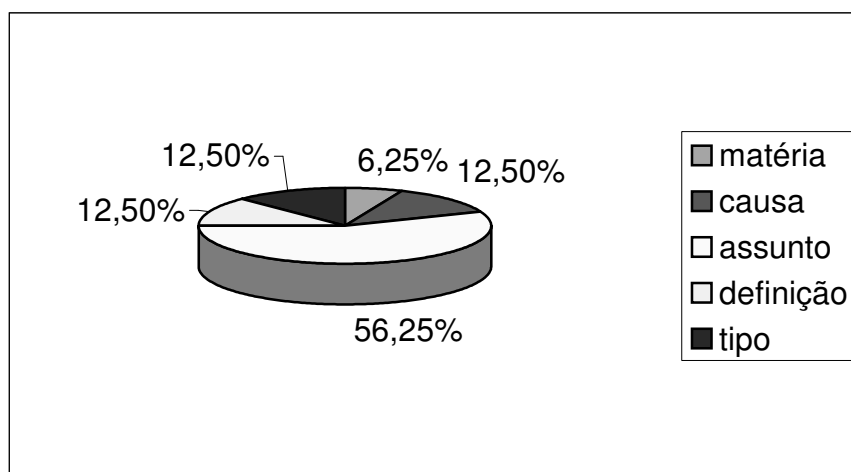
Quadro 29 - Comparativo da análise do valor semântico *assunto*.

Gráfico 10 - Porcentagens obtidas no enunciado 9.

O resultado da análise realizada pela pesquisadora indicando como *assunto* o valor semântico da preposição **de**, foi comprovado em 56,25%; 12,50% responderam que indica *causa*; 12,50% que indica *definição*; 12,50% que indica *tipo* ou a *qualidade*; e um caso (6,25%) que indica *matéria*.

Embora 56,25% de coincidências apontaram como valor semântico *assunto*, houve 5 discordâncias. No entanto, parece que respostas como “definição”, “tipo” e “qualidade” não representam valores semânticos da preposição **de** em “Fome de Amor”.

10. O DST é uma cifra **de** tempo para o veículo, é através dele que se calculado DSTR, que é uma medida da razão de performance da manufatura. [215]

Opções	Pesquisadora	Professores
origem (procedência de alguém ou de algo de algum lugar)		
posse (pertencer a alguém ou a alguma coisa)		
matéria (do que é feito)		
posição (onde está alguém ou alguma coisa)		
causa (por que uma coisa sucede)		
assunto ou objeto de que se trata		2
meio, instrumento ou modo		4
medida (tempo, espaço, volume, massa,...)	X	8
grupo ou série a que pertence		1
finalidade		
Outra. Qual? qualificação		1

Quadro 30 - Comparativo da análise do valor semântico *medida*.

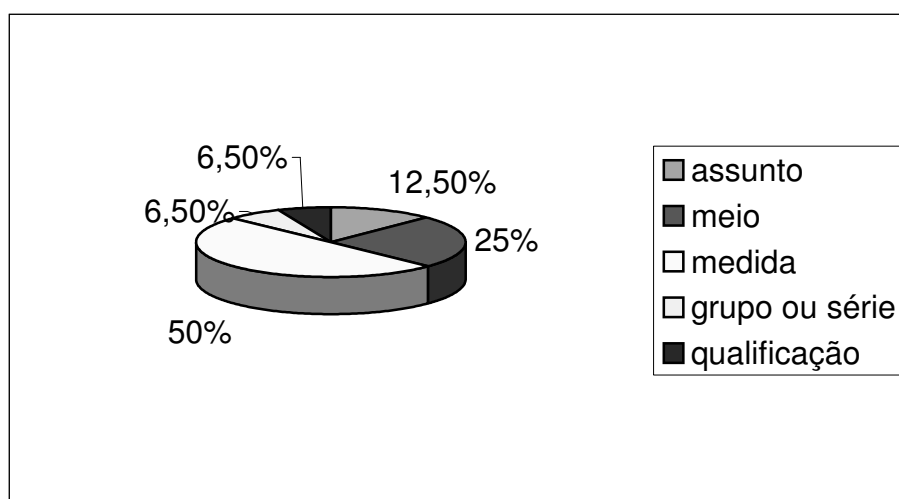


Gráfico 11 - Porcentagens obtidas no enunciado 10.

A comprovação da preposição **de** indicando *medida* foi de 50% em relação à análise da pesquisadora. 25% responderam que indica *meio, instrumento ou modo*; 12,50% que indica *assunto*; 6,50% que indica *grupo ou série a que pertence*; e 6,50% que indica *qualificação*.

Neste enunciado “cifra de tempo”, a resposta “qualificação” para a preposição **de** não apresenta valor semântico adequado.

11. A Fórmula 1 inicia hoje, no circuito **de** Nurburgring, na Alemanha, sua participação em terras européias na temporada 96. [16]

Opções	Pesquisadora	Professores
origem (procedência de alguém ou de algo de algum lugar)		
posse (pertencer a alguém ou a alguma coisa)	X	12
matéria (do que é feito)		
posição (onde está alguém ou alguma coisa)		1
causa (por que uma coisa sucede)		
assunto ou objeto de que se trata		
meio, instrumento ou modo		
medida (tempo, espaço, volume, massa,...)		
grupo ou série a que pertence		1
finalidade		
Outra. Qual? Espaço		2

Quadro 31 - Comparativo da análise do valor semântico *posse*.

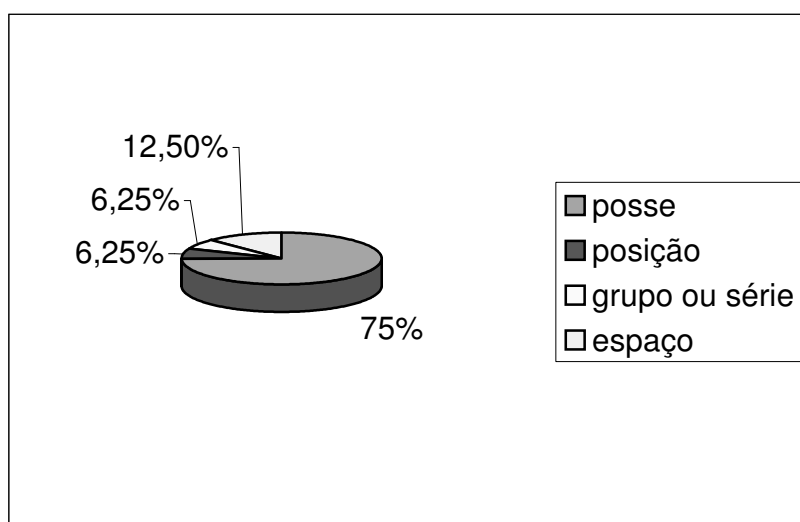


Gráfico 12 - Porcentagens obtidas no enunciado 11.

Neste enunciado, a análise realizada pela pesquisadora apresentou a idéia de posse, tendo o percentual de 75% nas respostas dos professores. 12,50% responderam que indica *espaço*; 6,25% que indica *posição*; e 6,25% que indica *grupo ou série a que pertence*.

12. O baú do dinheiro já foi aberto e, a partir de hoje, 11 das melhores Seleções masculinas **de** vôlei do planeta passam a procurar o caminho para chegar a ele. [73]

Opções	Pesquisadora	Professores
origem (procedência de alguém ou de algo de algum lugar)		
posse (pertencer a alguém ou a alguma coisa)		
matéria (do que é feito)		
posição (onde está alguém ou alguma coisa)		
causa (por que uma coisa sucede)		
assunto ou objeto de que se trata		1
meio, instrumento ou modo		
medida (tempo, espaço, volume, massa,...)		
grupo ou série a que pertence	X	15
finalidade		
Outra. Qual?		

Quadro 32 - Comparativo da análise do valor semântico *grupo ou série a que pertence*.

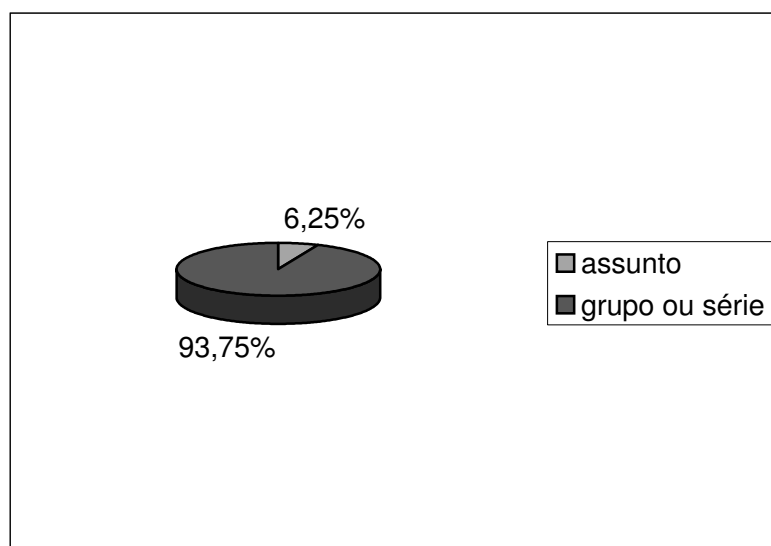


Gráfico 13 - Porcentagens obtidas no enunciado 12.

Houve comprovação de 93,75% com a análise realizada pela pesquisadora de que a preposição **de** indica *grupo ou série a que pertence*; e 6,25% (um caso) respondeu que indica *assunto*.

13. Para isso, será necessário projetar o componente e confeccionar novo molde para o processo **de** injeção de alumínio sob pressão. [197]

Opções	Pesquisadora	Professores
origem (procedência de alguém ou de algo de algum lugar)		
posse (pertencer a alguém ou a alguma coisa)		
matéria (do que é feito)		1
posição (onde está alguém ou alguma coisa)		
causa (por que uma coisa sucede)		1
assunto ou objeto de que se trata		2
meio, instrumento ou modo		6
medida (tempo, espaço, volume, massa,...)		
grupo ou série a que pertence		1
finalidade	X	5
Outra. Qual?		

Quadro 33 - Comparativo da análise do valor semântico *meio, instrumento ou modo*.

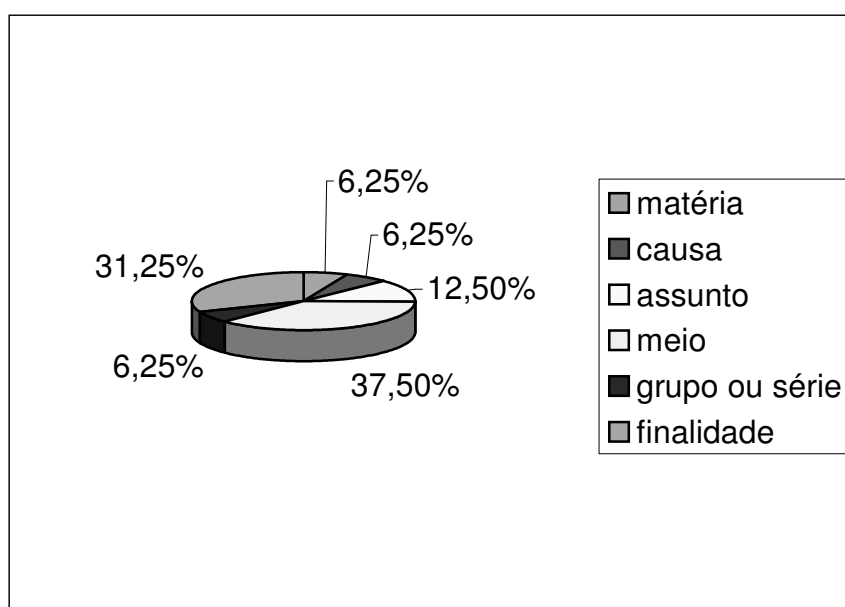


Gráfico 14 - Porcentagens obtidas no enunciado 13.

Houve comprovação de 31,25% com a análise realizada pela pesquisadora, ou seja, que indica *finalidade*. 37,50% responderam que indica *meio, instrumento ou modo*; 12,50% responderam que indica *assunto*; 6,25% que indica *matéria*; 6,25% que indica *causa*; e 6,25% que indica *grupo ou série a que pertence*.

Neste enunciado, o significado da preposição **de** apresentou grande diversidade, porém nenhuma alternativa superou os 40%.

14. Mercado ganha nova trava **de** volante. [313]

Opções	Pesquisadora	Professores
origem (procedência de alguém ou de algo de algum lugar)		
posse (pertencer a alguém ou a alguma coisa)		1
matéria (do que é feito)		
posição (onde está alguém ou alguma coisa)		
causa (por que uma coisa sucede)		
assunto ou objeto de que se trata		2
meio, instrumento ou modo		2
medida (tempo, espaço, volume, massa,...)		
grupo ou série a que pertence		
finalidade	X	10
Outra. Qual? tipificação		1

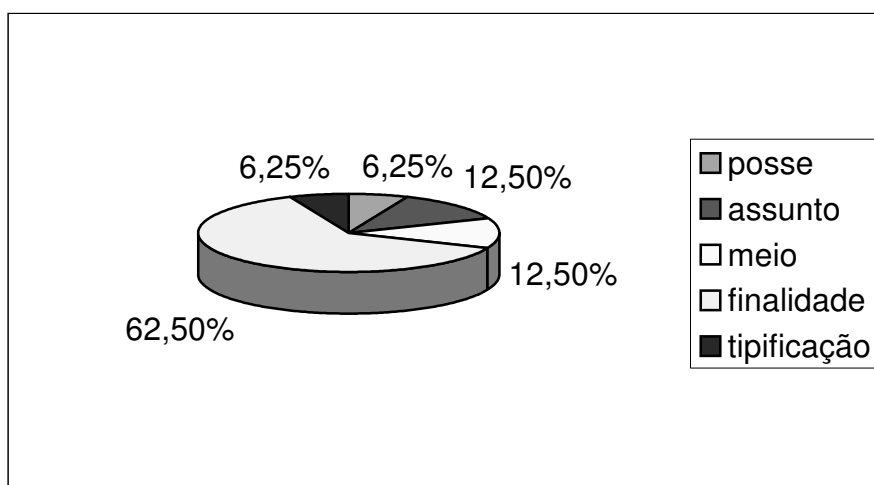
Quadro 34 - Comparativo da análise do valor semântico *finalidade*.

Gráfico 15 - Porcentagens obtidas no enunciado 14.

A análise realizada pela pesquisadora foi comprovada em 62,50%, ou seja, a preposição **de** indica *finalidade*. 12,50% responderam que indica *meio, instrumento ou modo*; outros 12,50% que indica *assunto*; 6,25% responderam que indica *posse*; e 6,25% que é uma *tipificação*.

15. ... técnicas da empresa transformadora ou de injeção das peças plásticas (conhecedores do processo de injeção **de** plásticos) e da ferramentaria responsável pela confecção dos moldes... [106]

Opções	Pesquisadora	Professores
origem (procedência de alguém ou de algo de algum lugar)		
posse (pertencer a alguém ou a alguma coisa)		
matéria (do que é feito)		7
posição (onde está alguém ou alguma coisa)		
causa (por que uma coisa sucede)		
assunto ou objeto de que se trata		4
meio, instrumento ou modo		
medida (tempo, espaço, volume, massa,...)		
grupo ou série a que pertence		3
finalidade	X	1
Outra. Qual? tipificação		1

Quadro 35 - Comparativo da análise do valor semântico *matéria*.

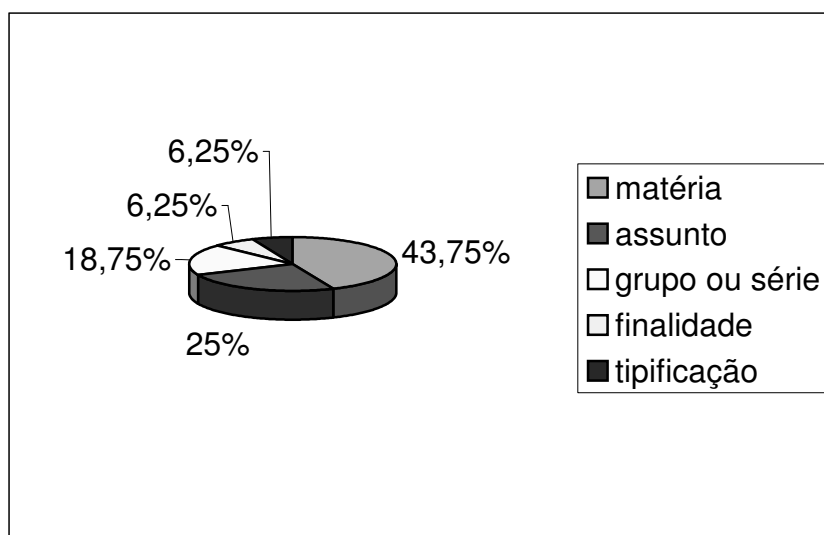


Gráfico 16 - Porcentagens obtidas no enunciado 15.

Somente 6,25% comprovaram a análise da pesquisadora de que a preposição **de** indica *finalidade*. 6,25% disseram que é uma *tipificação*; 18,75% que indica *grupo ou série a que pertence*; 25% que indica *assunto*; e 43,75% que indica *matéria*.

Neste enunciado (15) nenhuma alternativa ultrapassou o percentual de 50%.

O comparativo mostrou respostas bastante confiáveis em relação ao estudo da pesquisadora para a maioria dos enunciados, ou seja, enunciados 1, 2, 3, 4, 5, 6, 11, 12 e 14, correspondendo a 60% do total estudado. Em quatro enunciados (7, 8, 13 e 15) houve respostas muito divergentes, correspondendo a 26,6% do total dos enunciados. E, ainda, nos enunciados 9 e 10 (13,4%), as respostas coincidentes com as da pesquisadora ficaram na casa de 56,25% e 50% respectivamente, resultados ainda confiáveis.

5 CONCLUSÃO

Este trabalho mostra como os meios eletrônicos podem auxiliar nas pesquisas em linguagem natural. A Lingüística de *Corpus*, com o apoio de ferramentas informatizadas, pode contribuir de maneira eficaz na realização de análises lingüísticas, viabilizando a obtenção e exploração de dados empíricos e significativos da linguagem e possibilitando análises mais detalhadas. Ainda, o Processamento de Linguagem Natural permite a exploração desses dados para o desenvolvimento de estudos da Língua Portuguesa.

Com a utilização de um programa de análise de *corpus*, foram identificadas as palavras de maior centralidade na caracterização do gênero textual do *corpus* de estudo, composto pelos relatórios de estágio de alunos do Cefet-PR, para compará-lo com um *corpus* de referência, composto de textos compilados pelo NILC. O procedimento estatístico resultou na descoberta da preposição **de** como uma palavra central no *corpus* de estudo, embora alguns estudiosos a considerem apenas gramatical, possibilitando a realização do estudo semântico da referida preposição devido ao seu uso intensivo em sintagmas nominais.

O resultado da pesquisa, cujo foco foi determinar o significado da preposição **de**, quando esta se encontra na estrutura sintagmática N de N em enunciados extraídos aleatoriamente de um *corpus* que apresentou a preposição **de** como sendo uma palavra central em relatórios, pode ser observado e analisado no quadro 36.

Dos 15 enunciados selecionados aleatoriamente para o questionário em que se propunham alternativas de significação da preposição **de**, além de deixar para o respondente um campo aberto a outras possíveis, 73,4%, ou seja, 11 enunciados apresentaram as mesmas respostas em percentual superior a 50%, confirmando a

análise da pesquisadora; dentre eles dois enunciados (9 e 10) ficaram na casa dos 50% e 56,256% respectivamente. Já nos quatro enunciados restantes, houve respostas muito divergentes entre os próprios professores respondentes, apresentando percentuais inferiores a 30% em concordância com a pesquisadora.

Enunciados em que houve concordância com a pesquisadora		Enunciados em que não houve concordância com a pesquisadora	
Enunciado 1	100%	Enunciado 7	81,25%
Enunciado 2	68,75%	Enunciado 8	75%
Enunciado 3	68,75%	Enunciado 13	68,75%
Enunciado 4	93,75%	Enunciado 15	93,75%
Enunciado 5	62,50%		
Enunciado 6	62,50%		
Enunciado 9	56,25%		
Enunciado 10	50%		
Enunciado 11	75%		
Enunciado 12	93,75%		
Enunciado 14	62,50%		
11 enunciados.. 73,4%		4 enunciados26,6%	

Quadro 36 - Demonstrativo dos resultados

Pelo quadro acima, observa-se que em uma situação houve concordância plena em 100% com o que a apresentadora concluiu, ou seja, no enunciado 1... “liga de cobre” a preposição **de** carrega o significado de *matéria*.

Considerando que a preposição **de** é uma palavra gramatical, o estudo comprova que, além das preposições exercerem uma função morfosintática no enunciado, estas exercem papel importante no campo da semântica. Especificamente a preposição **de**, isolada não teria valor semântico, o que parece não acontecer com outras preposições, como com (junto, companhia), sob (embaixo), sobre (em cima). De outra forma, as palavras em geral, isoladamente, podem apresentar um (monossemia) ou mais significados (polissemia). No enunciado, via de regra, haverá apenas um valor semântico, salvo houver intenção de gerar ambigüidade (no caso da linguagem literária) ou ocorrer erro (anfibologia).

Assim, a preposição **de** no enunciado, embora pareça estar desprovida de significado e somente exercer a função de conectivo vocabular, ela pode apresentar significados diferentes, dependendo do contexto em que se encontra, conforme comprovado na análise semântica feita na pesquisa.

Além disso, observou-se, também, que na maioria dos casos, isto é, em 60% (1, 2, 3, 4, 5, 6, 11, 12 e 14) dos enunciados, houve concordância de 62%, ou acima desse percentual, nas alternativas assinaladas por parte dos respondentes, evidenciando que os significados escolhidos apresentam confiabilidade.

Com esta pesquisa, limitada ao estudo da preposição **de** no sintagma nominal N de N, a pesquisadora chegou aos seguintes significados: *matéria, origem, finalidade, assunto ou objeto de que se trata, medida, posse, grupo ou série a que pertence*. Isso não quer dizer que sejam apenas esses os valores semânticos desta preposição no referido sintagma, embora sejam os mais freqüentes em N de N.

Também pode-se afirmar que o valor semântico da preposição não fica restrito à preposição **de** estudada no sintagma citado N de N, pois, na Língua Portuguesa, todas as preposições têm significado quando no enunciado, e em qualquer outro sintagma em que se apresentem, seja no sintagma nominal ou verbal.

Outro aspecto a considerar é o pouco estudo existente em relação ao papel semântico desempenhado pela preposição no processo da comunicação, fato que parece não acontecer apenas na Língua Portuguesa, como também nas demais línguas neolatinas.

Isso, em parte, foi demonstrado no preenchimento do questionário, quando alguns professores respondentes encontraram dificuldade em assinalar a alternativa

correta ou “mais provável” de alguns enunciados, chegando inclusive a confundir semântica com morfologia e sintaxe.

Diante disso, espera-se que futuros trabalhos venham a ser feitos aprofundando o estudo da preposição **de**, não apenas no sintagma nominal estudado, mas também nos demais sintagmas, bem como outras pesquisas sejam desenvolvidas sobre o desempenho semântico de todas as preposições.

Fica como recomendação que no ensino da Língua se dê mais atenção à significação da preposição, sobretudo quando o assunto é regência nominal e verbal, não ficando apenas na “gramatiquice” repetindo de que se trata de um conectivo vocabular.

REFERÊNCIAS

ALLEN, J. **Natural language understanding**. Califórnia: Benjamin, Cummings Publishing Company, 1987.

ALMEIDA, N. M. de. **Gramática metódica da língua portuguesa**. São Paulo: Saraiva, 1998.

BAKHTIN, M. **Estética da criação verbal**. São Paulo: Martins Fontes, 2000.

BECHARA E. **Moderna gramática portuguesa**. Rio de Janeiro: Lucerna, 2003.

BIBER, D. *et al.* **Corpus linguistics - Investigating language structure and use**. Cambridge: Cambridge University Press, 1998.

COUTO, H. H. do. **Uma introdução à semiótica**. Rio de Janeiro: Presença, 1983.

CUNHA, C & CINTRA, L. **Nova gramática do português contemporâneo**. Rio de Janeiro: Nova Fronteira, 1985.

ECO, U. **O signo**. Lisboa: Presença, 1973.

FARACO, C. E & MOURA, F. M. de. **Gramática: fonética e fonologia, morfologia, sintaxe, estilística**. São Paulo: Ática, 1990.

FERREIRA, A. B. de H. **Novo Dicionário da Língua Portuguesa**. Rio de Janeiro: Nova Fronteira, 1986.

GARSIDE, R. *et al.* **Corpus annotation**. London & New York: Longman, 1997.

GERALDI, J.W. **Portos de passagem**. São Paulo: Martins Fontes, 2002.

HUNSTON, H. **Corpora in applied Linguistics**. New York: Cambridge University Press, 2002.

ILARI, R. & GERALDI, J. W. **Semântica**. São Paulo: Ática, 2002.

INFORMATIC. **História do computador**. Disponível em:
<http://www.informatic.hpg.ig.com.br/historia.htm>. Acesso em 5 de novembro de 2004.

KENNEDY, G. **An introduction to corpus linguistics**. London and New York: Longman, 1999.

KOCH, I. V. e SILVA, M. C. P. S. **Lingüística aplicada ao português: sintaxe**. São Paulo: Cortez, 1989.

KRIEGER, M^a da G. & FINATTO, M^a J. B. **Introdução à Terminologia: teoria & prática**. São Paulo: Contexto, 2004.

LIMA, C. H. da R. **Gramática normativa da língua portuguesa**. Rio de Janeiro: José Olympio, 2002.

LIMA, R. **Ciclo ginásial de português: primeira e segunda série**. Guanabara: Reper, 1968.

MACAMBIRA, J. R. **A estrutura morfo-sintática do Português**. São Paulo: Thomson, 1973.

MANGIAVACCHI, N. **Atividades de pesquisa**. Disponível em:
<<http://www.icmsc.sc.usp.br/norberto/geral/node23.htm>>. Acesso em 12 de maio de 2004.

MARTINS, D. S.; ZILBERKNOP, L. S. **Português instrumental**. Porto Alegre: Sagra-DC Luzzatto, 1993.

McENERY, T & WILSON, A. **Corpus linguistics**. Edinburgh: Edinburgh University Press, 2001.

MOURA, H. M. de M. **Significação e Contexto: uma introdução a questões de semântica e pragmática**. Florianópolis: Insular, 2000.

NUNES, Maria das Graças Volpe *et al.* **Introdução ao processamento das línguas naturais**. Notas Didáticas do ICMC, nº 38, São Carlos, junho de 1999.

OKA, C; ROPERTO, A. **Informática**, 2000. Disponível em:
<<http://www.cotianet.com.br/BIT/his/turing.htm>>. Acesso em 19 de julho de 2004.

PEREIRA, J. R; NAGAI, E. Y. **Os computadores podem pensar?** Disponível em:
<<http://www.din.uem.br/ia/maquinas/turing.htm>>. Acesso em 19 de julho de 2004.

Regulamento da Disciplina Estágio dos Cursos Superiores de Graduação do Cefet-PR

RUWET, N. **Introdução à Gramática Gerativa**. São Paulo: Perspectiva, 1967.

SARDINHA, T. B. **Artigo “Lingüística de Corpus: histórico e problemática”**. D.E.L.T.A., Vol. 16, Nº 2, 2000, p. 323-367.

_____ T. B. **Lingüística de Corpus**. São Paulo: Manoel, 2004.

SAUSSURE, F. de. **Curso de Lingüística Geral**. São Paulo: Cultrix,

VIEIRA, R.; LIMA, V. L. S. **Artigo “Lingüística computacional: princípios e aplicações”**. Porto Alegre (notas de aula).

ANEXO I